PCT

WORLD INTELLECTUAL PROPERTY ORGANIZATION International Bureau



INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification ⁶ : C07K 14/00	A2	(11) International Publication Number: WO 99/06439 (43) International Publication Date: 11 February 1999 (11.02.99)
(22) International Application Number: PCT/IBs (22) International Filing Date: 31 July 1998 (30) (30) Priority Data: 08/904,468 1 August 1997 (01.08.97) (71) Applicant (for all designated States except US): [FR/FR]; 24, rue Royale, F-75008 Paris (FR). (72) Inventors; and (75) Inventors/Applicants (for US only): DUMAS MIL WARDS, Jean-Baptiste [FR/FR]; 8, rue Grégoire of F-75006 Paris (FR). DUCLERT, Aymeric [FR/F] rue Victorine, F-94100 Saint-Maur (FR). LACROI [FR/FR]; 93, route de Vourles, F-69230 Saint-Ge (FR). (74) Agents: MARTIN, Jean-Jacques et al.; Cabinet Re 26, avenue Kléber, F-75116 Paris (FR).	GENSE NE EI de Tour R]; 6 te X, Brur nis Lav	BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GE, GH, GM, HR, HU, ID, IL, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, US, UZ, VN, YU, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG). Published Without international search report and to be republished upon receipt of that report.

(57) Abstract

The sequences of 5' ESTs derived from mRNAs encoding secreted proteins are disclosed. The 5' ESTs may be to obtain cDNAs and genomic DNAs corresponding to the 5' ESTs. The 5' ESTs may also be used in diagnostic, forensic, gene therapy, and chromosome mapping procedures. Upstream regulatory sequences may also be obtained using the 5' ESTs. The 5' ESTs may also be used to design expression vectors and secretion vectors.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav	TM	Turkmenistan
BF	Burkina Faso	GR	Greece		Republic of Macedonia	TR	Turkey
BG	Bulgaria	HU	Hungary	ML	Mali	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MN	Mongolia	ÜA	Ukraine
. BR	Brazil	ΙL	Israel	MR	Mauritania	UG	Uganda
BY	Belarus	IS	Iceland	MW	Malawi	US	United States of America
CA	Canada	lТ	Italy	MX	Mexico	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NE	Niger	VN	Viet Nam
CG	Congo	KE	Кепуа	NL	Netherlands	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NO	Norway	zw	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's	NZ	New Zealand		
CM	Cameroon		Republic of Korea	PL	Poland		
CN	China	KR	Republic of Korea	PT	Portugal		
CU	Cuba	KZ	Kazakstan	RO	Romania		
CZ	Czech Republic	LC	Saint Lucia	RU	Russian Federation		
DE	Germany	LI	Liechtenstein	SD	Sudan		
DK	Denmark	LK	Sri Lanka	SE	Sweden		
ER	Estonia .	LR	Liberia	SG	Singapore		

5

10

15

20

25

30

5' ESTs FOR SECRETED PROTEINS EXPRESSED IN ENDODERM

Background of the Invention

The estimated 50,000-100,000 genes scattered along the human chromosomes offer tremendous promise for the understanding, diagnosis, and treatment of human diseases. In addition, probes capable of specifically hybridizing to loci distributed throughout the human genome find applications in the construction of high resolution chromosome maps and in the identification of individuals.

In the past, the characterization of even a single human gene was a painstaking process, requiring years of effort. Recent developments in the areas of cloning vectors, DNA sequencing, and computer technology have merged to greatly accelerate the rate at which human genes can be isolated, sequenced, mapped, and characterized. Cloning vectors such as yeast artificial chromosomes (YACs) and bacterial artificial chromosomes (BACs) are able to accept DNA inserts ranging from 300 to 1000 kilobases (kb) or 100-400 kb in length respectively, thereby facilitating the manipulation and ordering of DNA sequences distributed over great distances on the human chromosomes. Automated DNA sequencing machines permit the rapid sequencing of human genes. Bioinformatics software enables the comparison of nucleic acid and protein sequences, thereby assisting in the characterization of human gene products.

Currently, two different approaches are being pursued for identifying and characterizing the genes distributed along the human genome. In one approach, large fragments of genomic DNA are isolated, cloned, and sequenced. Potential open reading frames in these genomic sequences are identified using bioinformatics software. However, this approach entails sequencing large stretches of human DNA which do not encode proteins in order to find the protein encoding sequences scattered throughout the genome. In addition to requiring extensive sequencing, the bioinformatics software may mischaracterize the genomic sequences obtained. Thus, the software may produce false positives in which noncoding DNA is mischaracterized as coding DNA or false negatives in which coding DNA is mischaracterized as non-coding DNA.

An alternative approach takes a more direct route to identifying and characterizing human genes. In this approach, complementary DNAs (cDNAs) are synthesized from isolated messenger RNAs (mRNAs) which encode human proteins. Using this approach,

sequencing is only performed on DNA which is derived from protein coding portions of the genome. Often, only short stretches of the cDNAs are sequenced to obtain sequences called expressed sequence tags (ESTs). The ESTs may then be used to isolate or purify extended cDNAs which include sequences adjacent to the EST sequences. The extended cDNAs may contain all of the sequence of the EST which was used to obtain them or only a portion of the sequence of the EST which was used to obtain them. In addition, the extended cDNAs may contain the full coding sequence of the gene from which the EST was derived or, alternatively, the extended cDNAs may include portions of the coding sequence of the gene from which the EST was derived. It will be appreciated that there may be several extended cDNAs which include the EST sequence as a result of alternate splicing or the activity of alternative promoters.

5

10

15

20

25

30

In the past, these short EST sequences were often obtained from oligo-dT primed cDNA libraries. Accordingly, they mainly corresponded to the 3' untranslated region of the mRNA. In part, the prevalence of EST sequences derived from the 3' end of the mRNA is a result of the fact that typical techniques for obtaining cDNAs are not well suited for isolating cDNA sequences derived from the 5' ends of mRNAs. (Adams et al., Nature 377:3-174, 1996; Hillier et al., Genome Res. 6:807-828, 1996).

In addition, in those reported instances where longer cDNA sequences have been obtained, the reported sequences typically correspond to coding sequences and do not include the full 5' untranslated region of the mRNA from which the cDNA is derived. Such incomplete sequences may not include the first exon of the mRNA, particularly in situations where the first exon is short. Furthermore, they may not include some exons, often short ones, which are located upstream of splicing sites. Thus, there is a need to obtain sequences derived from the 5' ends of mRNAs.

While many sequences derived from human chromosomes have practical applications, approaches based on the identification and characterization of those chromosomal sequences which encode a protein product are particularly relevant to diagnostic and therapeutic uses. Of the 50,000-100,000 protein coding genes, those genes encoding proteins which are secreted from the cell in which they are synthesized, as well as the secreted proteins themselves, are particularly valuable as potential therapeutic agents. Such proteins are often

3

involved in cell to cell communication and may be responsible for producing a clinically relevant response in their target cells.

In fact, several secretory proteins, including tissue plasminogen activator, G-CSF, GM-CSF, erythropoietin, human growth hormone, insulin, interferon-α, interferon-β, interferon-γ, and interleukin-2, are currently in clinical use. These proteins are used to treat a wide range of conditions, including acute myocardial infarction, acute ischemic stroke, anemia, diabetes, growth hormone deficiency, hepatitis, kidney carcinoma, chemotherapy induced neutropenia and multiple sclerosis. For these reasons, extended cDNAs encoding secreted proteins or portions thereof represent a particularly valuable source of therapeutic agents. Thus, there is a need for the identification and characterization of secreted proteins and the nucleic acids encoding them.

5

10

15

20

25

30

In addition to being therapeutically useful themselves, secretory proteins include short peptides, called signal peptides, at their amino termini which direct their secretion. These signal peptides are encoded by the signal sequences located at the 5' ends of the coding sequences of genes encoding secreted proteins. Because these signal peptides will direct the extracellular secretion of any protein to which they are operably linked, the signal sequences may be exploited to direct the efficient secretion of any protein by operably linking the signal sequences to a gene encoding the protein for which secretion is desired. In addition, portions of signal sequences may also be used to direct the intracellular import of a peptide or protein of interest. This may prove beneficial in gene therapy strategies in which it is desired to deliver a particular gene product to cells other than the cell in which it is produced. Signal sequences encoding signal peptides also find application in simplifying protein purification techniques. In such applications, the extracellular secretion of the desired protein greatly facilitates purification by reducing the number of undesired proteins from which the desired protein must be selected. Thus, there exists a need to identify and characterize the 5' portions of the genes for secretory proteins which encode signal peptides.

Public information on the number of human genes for which the promoters and upstream regulatory regions have been identified and characterized is quite limited. In part, this may be due to the difficulty of isolating such regulatory sequences. Upstream regulatory sequences such as transcription factor binding sites are typically too short to be utilized as probes for isolating promoters from human genomic libraries. Recently, some approaches

4

have been developed to isolate human promoters. One of them consists of making a CpG island library (Cross, et al., Nature Genetics 6: 236-244, 1994). The second consists of isolating human genomic DNA sequences containing SpeI binding sites by the use of SpeI binding protein. (Mortlock et al., Genome Res. 6:327-335, 1996). Both of these approaches have their limits due to a lack of specificity or of comprehensiveness.

5

10

15

20

25

30

The present 5' ESTs may be used to efficiently identify and isolate upstream regulatory regions which control the location, developmental stage, rate, and quantity of protein synthesis, as well as the stability of the mRNA. (Theil, *BioFactors* 4:87-93, 1993). Once identified and characterized, these regulatory regions may be utilized in gene therapy or protein purification schemes to obtain the desired amount and locations of protein synthesis or to inhibit, reduce, or prevent the synthesis of undesirable gene products.

In addition, ESTs containing the 5' ends of secretory protein genes may include sequences useful as probes for chromosome mapping and the identification of individuals. Thus, there is a need to identify and characterize the sequences upstream of the 5' coding sequences of genes encoding secretory proteins.

Summary of the Invention

The present invention relates to purified, isolated, or recombinant ESTs which include sequences derived from the authentic 5' ends of their corresponding mRNAs. The term "corresponding mRNA" refers to the mRNA which was the template for the cDNA synthesis which produced the 5' EST. These sequences will be referred to hereinafter as "5' ESTs." As used herein, the term "purified" does not require absolute purity; rather, it is intended as a relative definition. Individual 5' EST clones isolated from a cDNA library have been conventionally purified to electrophoretic homogeneity. The sequences obtained from these clones could not be obtained directly either from the library or from total human DNA. The cDNA clones are not naturally occurring as such, but rather are obtained via manipulation of a partially purified naturally occurring substance (messenger RNA). The conversion of mRNA into a cDNA library involves the creation of a synthetic substance (cDNA) and pure individual cDNA clones can be isolated from the synthetic library by clonal selection. Thus, creating a cDNA library from messenger RNA and subsequently isolating individual clones from that library results in an approximately 10⁴-10⁶ fold purification of the native message.

5

10

15

20

25

30

Purification of starting material or natural material to at least one order of magnitude, preferably two or three orders, and more preferably four or five orders of magnitude is expressly contemplated.

As used herein, the term "isolated" requires that the material be removed from its original environment (e.g., the natural environment if it is naturally occurring). For example, a naturally-occurring polynucleotide present in a living animal is not isolated, but the same polynucleotide, separated from some or all of the coexisting materials in the natural system, is isolated.

As used herein, the term "recombinant" means that the 5' EST is adjacent to "backbone" nucleic acid to which it is not adjacent in its natural environment. Additionally, to be "enriched" the 5' ESTs will represent 5% or more of the number of nucleic acid inserts in a population of nucleic acid backbone molecules. Backbone molecules according to the present invention include nucleic acids such as expression vectors, self-replicating nucleic acids, viruses, integrating nucleic acids, and other vectors or nucleic acids used to maintain or manipulate a nucleic acid insert of interest. Preferably, the enriched 5' ESTs represent 15% or more of the number of nucleic acid inserts in the population of recombinant backbone molecules. More preferably, the enriched 5' ESTs represent 50% or more of the number of nucleic acid inserts in the population of recombinant backbone molecules. In a highly preferred embodiment, the enriched 5' ESTs represent 90% or more of the number of nucleic acid inserts in the population of recombinant backbone molecules.

"Stringent", moderate," and "low" hybridization conditions are as defined in Example 29.

Unless otherwise indicated, a "complementary" sequence is fully complementary.

Thus, 5' ESTs in cDNA libraries in which one or more 5' ESTs make up 5% or more of the number of nucleic acid inserts in the backbone molecules are "enriched recombinant 5' ESTs" as defined herein. Likewise, 5' ESTs in a population of plasmids in which one or more 5' EST of the present invention have been inserted such that they represent 5% or more of the number of inserts in the plasmid backbone are "enriched recombinant 5' ESTs" as defined herein. However, 5' ESTs in cDNA libraries in which 5' ESTs constitute less than 5% of the number of nucleic acid inserts in the population of backbone molecules, such as libraries in

6

which backbone molecules having a 5' EST insert are extremely rare, are not "enriched recombinant 5' ESTs."

In particular, the present invention relates to 5' ESTs which are derived from genes encoding secreted proteins. As used herein, a "secreted" protein is one which, when expressed in a suitable host cell, is transported across or through a membrane, including transport as a result of signal peptides in its amino acid sequence. "Secreted" proteins include without limitation proteins secreted wholly (e.g. soluble proteins), or partially (e.g. receptors) from the cell in which they are expressed. "Secreted" proteins also include without limitation proteins which are transported across the membrane of the endoplasmic reticulum.

5

10

15

20

25

30

Such 5' ESTs include nucleic acid sequences, called signal sequences, which encode signal peptides which direct the extracellular secretion of the proteins encoded by the genes from which the 5' ESTs are derived. Generally, the signal peptides are located at the amino termini of secreted proteins.

Secreted proteins are translated by ribosomes associated with the "rough" endoplasmic reticulum. Generally, secreted proteins are co-translationally transferred to the membrane of the endoplasmic reticulum. Association of the ribosome with the endoplasmic reticulum during translation of secreted proteins is mediated by the signal peptide. The signal peptide is typically cleaved following its co-translational entry into the endoplasmic reticulum. After delivery to the endoplasmic reticulum, secreted proteins may proceed through the Golgi apparatus. In the Golgi apparatus, the proteins may undergo post-translational modification before entering secretory vesicles which transport them across the cell membrane.

The 5' ESTs of the present invention have several important applications. For example, they may be used to obtain and express cDNA clones which include the full protein coding sequences of the corresponding gene products, including the authentic translation start sites derived from the 5' ends of the coding sequences of the mRNAs from which the 5' ESTs are derived. These cDNAs will be referred to hereinafter as "full length cDNAs." These cDNAs may also include DNA derived from mRNA sequences upstream of the translation start site. The full length cDNA sequences may be used to express the proteins corresponding to the 5' ESTs. As discussed above, secreted proteins are therapeutically important. Thus, the proteins expressed from the cDNAs may be useful in treating or

7

controlling a variety of human conditions. The 5' ESTs may also be used to obtain the corresponding genomic DNA. The term "corresponding genomic DNA" refers to the genomic DNA which encodes the mRNA from which the 5' EST was derived.

5

10

15

20

25

30

Alternatively, the 5' ESTs may be used to obtain and express extended cDNAs encoding portions of the secreted protein. The portions may comprise the signal peptides of the secreted proteins or the mature proteins generated when the signal peptide is cleaved off. The portions may also comprise polypeptides having at least 10 consecutive amino acids encoded by the extended cDNAs or full length cDNAs. Alternatively, the portions may comprise at least 15 consecutive amino acids encoded by the extended cDNAs or full length cDNAs. In some embodiments, the portions may comprise at least 25 consecutive amino acids encoded by the extended cDNAs or full length cDNAs. In other embodiments, the portions may comprise at least 40 amino acids encoded by the extended cDNAs or full length cDNAs.

Antibodies which specifically recognize the entire secreted proteins encoded by the extended cDNAs, full length cDNAs, or fragments thereof having at least 10 consecutive amino acids, at least 15 consecutive amino acids, at least 25 consecutive amino acids, or at least 40 consecutive amino acids may also be obtained as described below. Antibodies which specifically recognize the mature protein generated when the signal peptide is cleaved may also be obtained as described below. Similarly, antibodies which specifically recognize the signal peptides encoded by the extended cDNAs or full length cDNAs may also be obtained.

In some embodiments, the extended cDNAs obtained using the 5' ESTs include the signal sequence. In other embodiments, the extended cDNAs obtained using the 5' ESTs may include the full coding sequence for the mature protein (i.e. the protein generated when the signal polypeptide is cleaved off). In addition, the extended cDNAs obtained using the 5' ESTs may include regulatory regions upstream of the translation start site or downstream of the stop codon which control the amount, location, or developmental stage of gene expression.

As discussed above, secreted proteins are therapeutically important. Thus, the proteins expressed from the extended cDNAs or full length cDNAs obtained using the 5' ESTs may be useful in treating or controlling a variety of human conditions.

8

The 5' ESTs (or cDNAs or genomic DNAs obtained therefrom) may be used in forensic procedures to identify individuals or in diagnostic procedures to identify individuals having genetic diseases resulting from abnormal expression of the genes corresponding to the 5' ESTs. In addition, the present invention is useful for constructing a high resolution map of the human chromosomes.

5

10

15

20

25

The present invention also relates to secretion vectors capable of directing the secretion of a protein of interest. Such vectors may be used in gene therapy strategies in which it is desired to produce a gene product in one cell which is to be delivered to another location in the body. Secretion vectors may also facilitate the purification of desired proteins.

The present invention also relates to expression vectors capable of directing the expression of an inserted gene in a desired spatial or temporal manner or at a desired level. Such vectors may include sequences upstream of the 5' ESTs, such as promoters or upstream regulatory sequences.

Finally, the present invention may also be used for gene therapy to control or treat genetic diseases. Signal peptides may also be fused to heterologous proteins to direct their extracellular secretion.

Bacterial clones containing Bluescript plasmids having inserts containing the 5' ESTs of the present invention (SEQ ID NOs: 38-184 are presently stored at 80°C in 4% (v/v) glycerol in the inventor's laboratories under the designations listed next to the SEQ ID NOs in II). The inserts may be recovered from the deposited materials by growing the appropriate clones on a suitable medium. The Bluescript DNA can then be isolated using plasmid isolation procedures familiar to those skilled in the art such as alkaline lysis minipreps or large scale alkaline lysis plasmid isolation procedures. If desired the plasmid DNA may be further enriched by centrifugation on a cesium chloride gradient, size exclusion chromatography, or anion exchange chromatography. The plasmid DNA obtained using these procedures may then be manipulated using standard cloning techniques familiar to those skilled in the art. Alternatively, a PCR can be done with primers designed at both ends of the EST insertion. The PCR product which corresponds to the 5' EST can then be manipulated using standard cloning techniques familiar to those skilled in the art.

One aspect of the present invention is a purified or isolated nucleic acid having the sequence of one of SEQ ID NOs: 38-184 or having a sequence complementary thereto. In one embodiment, the nucleic acid is recombinant.

Another aspect of the present invention is a purified or isolated nucleic acid comprising at least 10 consecutive bases of the sequence of one of SEQ ID NOs: 38-184 or one of the sequences complementary thereto.

5

10

15

20

25

30

Yet another aspect of the present invention is a purified or isolated nucleic acid comprising at least 15 consecutive bases of one of the sequences of SEQ ID NOs: 38-184 or one of the sequences complementary thereto. In one embodiment, the nucleic acid is recombinant.

A further aspect of the present invention is a purified or isolated nucleic acid of at least 15 bases capable of hybridizing under stringent conditions to the sequence of one of SEQ ID NOs: 38-184 or one of the sequences complementary to the sequences of SEQ ID NOs: 38-184. In one embodiment, the nucleic acid is recombinant.

Another aspect of the present invention is a purified or isolated nucleic acid encoding a human gene product, said human gene product having a sequence partially encoded by one of the sequences of SEQ ID NO: 38-184.

Still another aspect of the present invention is a method of making a cDNA encoding a human secretory protein, said human secretory protein being partially encoded by one of SEQ ID NOs 38-184, comprising the steps of contacting a collection of mRNA molecules from human cells with a primer comprising at least 15 consecutive nucleotides of a sequence complementary to one of SEQ ID NOs: 38-184; hybridizing said primer to an mRNA in said collection that encodes said protein; reverse transcribing said hybridized primer to make a first cDNA strand from said mRNA; making a second cDNA strand complementary to said first cDNA strand; and isolating the resulting cDNA encoding said protein comprising said first cDNA strand and said second cDNA strand.

Another aspect of the invention is an isolated or purified cDNA encoding a human secretory protein, said human secretory protein comprising the protein encoded by one of SEQ ID NOs 38-184 or a fragment thereof of at least 10 amino acids, said cDNA being obtainable by the method described in the preceding paragraph. In one embodiment, the

10

cDNA comprises the full protein coding sequence of said protein which sequence is partially included in one of the sequences of SEQ ID NOs: 38-184.

Another aspect of the present invention is a method of making a cDNA encoding a human secretory protein that is partially encoded by one of SEQ ID NOs 38-184, comprising the steps of obtaining a cDNA comprising one of the sequences of SEQ ID NOs: 38-184; contacting said cDNA with a detectable probe comprising at least 15 consecutive nucleotides of said sequence of SEQ ID NO: 38-184 or a sequence complementary thereto under conditions which permit said probe to hybridize to said cDNA; identifying a cDNA which hybridizes to said detectable probe; and isolating said cDNA which hybridizes to said probe.

5

10

15

20

25

30

Another aspect of the present invention is an isolated or purified cDNA encoding a human secretory protein, said human secretory protein comprising the protein encoded by one of SEQ ID NOs 38-184 or a fragment thereof of at least 10 amino acids, said cDNA being obtainable by the method described in the preceding paragraph. In one embodiment, the cDNA comprises the full protein coding sequence partially included in one of the sequences of SEQ ID NOs: 38-184.

Another aspect of the present invention is a method of making a cDNA comprising one of the sequence of SEQ ID NOs: 38-184, comprising the steps of contacting a collection of mRNA molecules from human cells with a first primer capable of hybridizing to the polyA tail of said mRNA; hybridizing said first primer to said polyA tail; reverse transcribing said mRNA to make a first cDNA strand; making a second cDNA strand complementary to said first cDNA strand using at least one primer comprising at least 15 nucleotides of one of the sequences of SEQ ID NOs 38-184; and isolating the resulting cDNA comprising said first cDNA strand and said second cDNA strand.

Another aspect of the present invention is an isolated or purified cDNA encoding a human secretory protein, said human secretory protein comprising the protein encoded by one of SEQ ID NOs 38-184 or a fragment thereof of at least 10 amino acids, said cDNA being obtainable by the method described in the preceding paragraph. In one embodiment, the cDNA comprises the full protein coding sequence partially included in one of the sequences of SEQ ID NOs: 38-184.

In one embodiment of the method described in the two paragraphs above, the second cDNA strand is made by contacting said first cDNA strand with a first pair of primers, said

11

first pair of primers comprising a second primer comprising at least 15 consecutive nucleotides of one of the sequences of SEQ ID NOs 38-184 and a third primer having a sequence therein which is included within the sequence of said first primer, performing a first polymerase chain reaction with said first pair of nested primers to generate a first PCR product; contacting said first PCR product with a second pair of primers, said second pair of primers comprising a fourth primer, said fourth primer comprising at least 15 consecutive nucleotides of said sequence of one of SEQ ID NOs: 38-184, and a fifth primer, said fourth and fifth primers being capable of hybridizing to sequences within said first PCR product; and performing a second polymerase chain reaction, thereby generating a second PCR product.

10

5

One aspect of the present invention is an isolated or purified cDNA encoding a human secretory protein, said human secretory protein comprising the protein encoded by one of SEQ ID NOs 38-184, or a fragment thereof of at least 10 amino acids, said cDNA being obtainable by the method of the preceding paragraph. In one embodiment, the cDNA comprises the full protein coding sequence partially included in one of the sequences of SEQ ID NOs: 38-184.

15

Another aspect of the present invention is the method described four paragraphs above in which the second cDNA strand is made by contacting said first cDNA strand with a second primer comprising at least 15 consecutive nucleotides of the sequences of SEQ ID NOs: 38-184; hybridizing said second primer to said first strand cDNA; and extending said hybridized second primer to generate said second cDNA strand.

20

Another aspect of the present invention is an isolated or purified cDNA encoding a human secretory protein, said human secretory protein comprising the protein partially encoded by one of SEQ ID NOs 38-184 or comprising a fragment thereof of at least 10 amino acids, said cDNA being obtainable by the method described in the preceding paragraph. In one embodiment, the cDNA comprises the full protein coding sequence partially included in of one of the sequences of SEQ ID NOs: 38-184.

25

30

Another aspect of the present invention is a method of making a protein comprising one of the sequences of SEQ ID NOs: 185-331, comprising the steps of obtaining a cDNA encoding the full protein sequence partially included in one of the sequences of sequence of SEQ ID NOs: 38-184; inserting said cDNA in an expression vector such that said cDNA is

operably linked to a promoter; introducing said expression vector into a host cell whereby said host cell produces the protein encoded by said cDNA; and isolating said protein.

Another aspect of the present invention is an isolated protein obtainable by the method described in the preceding paragraph.

Another aspect of the present invention is a method of obtaining a promoter DNA comprising the steps of obtaining DNAs located upstream of the nucleic acids of SEQ ID NOs: 38-184 or the sequences complementary thereto; screening said upstream DNAs to identify a promoter capable of directing transcription initiation; and isolating said DNA comprising said identified promoter. In one embodiment, the obtaining step comprises chromosome walking from said nucleic acids of SEQ ID NOs: 38-184 or sequences complementary thereto. In another embodiment, the screening step comprises inserting said upstream sequences into a promoter reporter vector. In another embodiment, the screening step comprises identifying motifs in said upstream DNAs which are transcription factor binding sites or transcription start sites.

Another aspect of the present invention is an isolated promoter obtainable by the method described above.

Another aspect of the present invention is an isolated or purified protein comprising one of the sequences of SEQ ID NOs: 185-331.

Another aspect of the present invention is the inclusion of at least one of the sequences of SEQ ID NOs: 38-184, or one of the sequences complementary to the sequences of SEQ ID NOs: 38-184, or a fragment thereof of at least 15 consecutive nucleotides in an array of discrete ESTs or fragments thereof of at least 15 nucleotides in length. In one embodiment, the array includes at least two of the sequences of SEQ ID NOs: 38-184, the sequences complementary to the sequences of SEQ ID NOs: 38-184, or fragments thereof of at least 15 consecutive nucleotides. In another embodiment, the array includes at least five of the sequences of SEQ ID NOs: 38-184, the sequences complementary to the sequences of SEQ ID NOs: 38-184, or fragments thereof of at least 15 consecutive nucleotides.

Another aspect of the present invention is a promoter having a sequence selected from the group consisting of SEQ ID NOs: 31, 34, and 37.

5

10

15

20

25

13

Brief Description of the Drawings

Figure 1 is a summary of a procedure for obtaining cDNAs which have been selected to include the 5' ends of the mRNAs from which they derived.

Figure 2 shows the distribution of Von Heijne scores for 5' ESTs in each of the categories described herein and the probability that these 5' ESTs encode a signal peptide.

5

10

15

25

30

Figure 3 summarizes a general method used to clone and sequence extended cDNAs containing sequences adjacent to 5' ESTs.

Figure 4 (description of promoters structure isolated from SignalTag 5' ESTs) provides a schematic description of promoters isolated and the way they are assembled with the corresponding 5' tags.

Detailed Description of the Preferred Embodiment

Table IV is an analysis of the 43 amino acids located at the N terminus of all human SwissProt proteins to determine the frequency of false positives and false negatives using the techniques for signal peptide identification described herein.

Table V shows the distribution of 5' ESTs in each category described herein and the number of 5' ESTs in each category having a given minimum Von Heijne's score.

Table VI shows the distribution of 5' ESTs in each category described herein with respect to the tissue from which the 5' ESTs of the corresponding mRNA were obtained.

Table VII describes the transcription factor binding sites present in each of these 20 promoters.

I. General Methods for Obtaining 5' ESTs derived from mRNAs with intact 5' ends

In order to obtain the 5' ESTs of the present invention, mRNAs with intact 5' ends must be obtained. Currently, there are two approaches for obtaining such mRNAs with intact 5' ends as described below: either chemical (1) or enzymatic (2).

1. Chemical Methods for Obtaining mRNAs having Intact 5' Ends

One of these approaches is a chemical modification method involving derivatization of the 5' ends of the mRNAs and selection of the derivatized mRNAs. The 5' ends of eukaryotic mRNAs possess a structure referred to as a "cap" which comprises a guanosine

methylated at the 7 position. The cap is joined to the first transcribed base of the mRNA by a 5′, 5′-triphosphate bond. In some instances, the 5′ guanosine is methylated in both the 2 and 7 positions. Rarely, the 5′ guanosine is trimethylated at the 2, 7 and 7 positions. In the chemical method for obtaining mRNAs having intact 5′ ends, the 5′ cap is specifically derivatized and coupled to a reactive group on an immobilizing substrate. This specific derivatization is based on the fact that only the ribose linked to the methylated guanosine at the 5′ end of the mRNA and the ribose linked to the base at the 3′ terminus of the mRNA, possess 2′, 3′-cis diols.

Optionally, the 2', 3'-cis diol of the 3' terminal ribose may be chemically modified, substituted, converted, or eliminated, leaving only the ribose linked to the methylated guanosine at the 5' end of the mRNA with a 2', 3'-cis diol. A variety of techniques are available for eliminating the 2', 3'-cis diol on the 3' terminal ribose. For example, controlled alkaline hydrolysis may be used to generate mRNA fragments in which the 3' terminal ribose is a 3'-phosphate, 2'-phosphate or (2', 3')-cyclophosphate. Thereafter, the fragment which includes the original 3' ribose may be eliminated from the mixture through chromatography on an oligodT column. Alternatively, a base which lacks the 2', 3'-cis diol may be added to the 3' end of the mRNA using an RNA ligase such as T4 RNA ligase. Example 1 below describes a method for ligation of a nucleoside diphosphate to the 3' end of messenger RNA.

20

25

5

10

15

EXAMPLE 1

Ligation of the Nucleoside Diphosphate pCp to the 3' End of mRNA.

One µg of RNA was incubated in a final reaction medium of 10 µl in the presence of 5 U of T₄ phage RNA ligase in the buffer provided by the manufacturer (Gibco - BRL), 40 U of the RNase inhibitor RNasin (Promega) and, 2 µl of ³²pCp (Amersham #PB 10208). The incubation was performed at 37°C for 2 hours or overnight at 7-8°C.

Following modification or elimination of the 2', 3'-cis diol at the 3' ribose, the 2', 3'-cis diol present at the 5' end of the mRNA may be oxidized using reagents such as NaBH, NaBH, CN, or sodium periodate, thereby converting the 2', 3'-cis diol to a dialdehyde.

Example 2 describes the oxidation of the 2', 3'-cis diol at the 5' end of the mRNA with sodium periodate.

EXAMPLE 2

Oxidation of 2', 3'-cis diol at the 5' End of the mRNA with Sodium Periodate

0.1 OD unit of either a capped oligoribonucleotide of 47 nucleotides (including the cap) or an uncapped oligoribonucleotide of 46 nucleotides were treated as follows. The oligoribonucleotides were produced by *in vitro* transcription using the transcription kit "AmpliScribe T7" (Epicentre Technologies). As indicated below, the DNA template for the RNA transcript contained a single cytosine. To synthesize the uncapped RNA, all four NTPs were included in the *in vitro* transcription reaction. To obtain the capped RNA, GTP was replaced by an analogue of the cap, m7G(5')ppp(5')G. This compound, recognized by the polymerase, was incorporated into the 5' end of the nascent transcript during the initiation of transcription but was not incorporated during the extension step. Consequently, the resulting RNA contained a cap at its 5' end. The sequences of the oligoribonucleotides produced by the *in vitro* transcription reaction were:

+Cap:

5'm7GpppGCAUCCUACUCCCAUCCAAUUCCACCCUAACUCCUCCAUCUCCAC-3' (SEQ ID NO:1)

20 -Cap:

25

30

5

10

15

5'-pppGCAUCCUACUCCAUCCAAUUCCACCCUAACUCCUCCCAUCUCCAC-3' (SEQ ID NO:2)

The oligoribonucleotides were dissolved in 9 μ l of acetate buffer (0.1 M sodium acetate, pH 5.2) and 3 μ l of freshly prepared 0.1 M sodium periodate solution. The mixture was incubated for 1 hour in the dark at 4°C or room temperature. Thereafter, the reaction was stopped by adding 4 μ l of 10% ethylene glycol. The product was ethanol precipitated, resuspended in at least 10 μ l of water or appropriate buffer and dialyzed against water.

The resulting aldehyde groups may then be coupled to molecules having a reactive amine group, such as hydrazine, carbazide, thiocarbazide or semicarbazide groups, in order to facilitate enrichment of the 5' ends of the mRNAs. Molecules having reactive amine groups

which are suitable for use in selecting mRNAs having intact 5' ends include avidin, proteins, antibodies, vitamins, ligands capable of specifically binding to receptor molecules, or oligonucleotides. Example 3 below describes the coupling of the resulting dialdehyde to biotin.

5

10

15

20

25

EXAMPLE 3

Coupling of the Dialdehyde at the 5' End of Transcripts with Biotin

The oxidation product obtained in Example 2 was dissolved in 50 μ l of sodium acetate at a pH between 5 and 5.2 and 50 μ l of freshly prepared 0.02 M solution of biotin hydrazide in a methoxyethanol/water mixture (1:1) of formula:

EXAMPLE 4

Specificity of Biotinylation of Capped Transcripts

The specificity of the biotinylation for capped mRNAs was evaluated by gel electrophoresis of the following samples:

Sample 1. The 46 nucleotide uncapped *in vitro* transcript prepared as in Example 2 and labeled with ³²pCp as described in Example 1.

Sample 2. The 46 nucleotide uncapped *in vitro* transcript prepared as in Example 2, labeled with ³²pCp as described in Example 1, treated with the oxidation reaction of Example 2, and subjected to the biotinylation conditions of Example 3.

Sample 3. The 47 nucleotide capped *in vitro* transcript prepared as in Example 2 and labeled with ³²pCp as described in Example 1.

Sample 4. The 47 nucleotide capped *in vitro* transcript prepared as in Example 2, labeled with ³²pCp as described in Example 1, treated with the oxidation reaction of Example 2, and subjected to the biotinylation conditions of Example 3.

17

Samples 1 and 2 had identical migration rates, demonstrating that the uncapped RNAs were not oxidized and biotinylated. Sample 3 migrated more slowly than Samples 1 and 2, while Sample 4 exhibited the slowest migration. The difference in migration of the RNAs in Samples 3 and 4 demonstrates that the capped RNAs were specifically biotinylated.

5

In some cases, mRNAs having intact 5' ends may be enriched by binding the molecule containing a reactive amine group to a suitable solid phase substrate such as the inside of the vessel containing the mRNAs, magnetic beads, chromatography matrices, or nylon or nitrocellulose membranes. For example, where the molecule having a reactive amine group is biotin, the solid phase substrate may be coupled to avidin or streptavidin. Alternatively, where the molecule having the reactive amine group is an antibody or receptor ligand, the solid phase substrate may be coupled to the cognate antigen or receptor. Finally, where the molecule having a reactive amine group comprises an oligonucleotide, the solid phase substrate may comprise a complementary oligonucleotide.

15

20

10

The mRNAs having intact 5' ends may be released from the solid phase following the enrichment procedure. For example, where the dialdehyde is coupled to biotin hydrazide and the solid phase comprises streptavidin, the mRNAs may be released from the solid phase by simply heating to 95 degrees Celsius in 2% SDS. In some methods, the molecule having a reactive amine group may also be cleaved from the mRNAs having intact 5' ends following enrichment. Example 5 describes the capture of biotinylated mRNAs with streptavidin coated beads and the release of the biotinylated mRNAs from the beads following enrichment.

EXAMPLE 5

Capture and Release of Biotinylated mRNAs Using Streptavidin Coated Beads

25

30

The streptavidin coated magnetic beads were prepared according to the manufacturer's instructions (CPG Inc., USA). The biotinylated mRNAs were added to a hybridization buffer (1.5 M NaCl, pH 5 - 6). After incubating for 30 minutes, the unbound and nonbiotinylated material was removed. The beads were then washed several times in water with 1% SDS. The beads thus obtained were incubated for 15 minutes at 95°C in water containing 2% SDS.

18

Example 6 demonstrates the efficiency with which biotinylated mRNAs were recovered from the streptavidin coated beads.

EXAMPLE 6

5

Efficiency of Recovery of Biotinylated mRNAs

The efficiency of the recovery procedure was evaluated as follows. Capped RNAs were labeled with ³²pCp, oxidized, biotinylated and bound to streptavidin coated beads as described above. Subsequently, the bound RNAs were incubated for 5, 15 or 30 minutes at 95°C in the presence of 2% SDS.

10

The products of the reaction were analyzed by electrophoresis on 12% polyacrylamide gels under denaturing conditions (7 M urea). The gels were subjected to autoradiography. During this manipulation, the hydrazone bonds were not reduced.

Increasing amounts of nucleic acids were recovered as incubation times in 2% SDS increased, demonstrating that biotinylated mRNAs were efficiently recovered.

15

In an alternative method for obtaining mRNAs having intact 5' ends, an oligonucleotide which has been derivatized to contain a reactive amine group is specifically coupled to mRNAs having an intact cap. Preferably, the 3' end of the mRNA is blocked prior to the step in which the aldehyde groups are joined to the derivatized oligonucleotide, as described above, so as to prevent the derivatized oligonucleotide from being joined to the 3' end of the mRNA using T4 RNA ligase as described in example 1. However, as discussed above, blocking the 3' end of the mRNA is an optional step. Derivatized oligonucleotides may be prepared as described in Example 7.

25

30

20

EXAMPLE 7

Derivatization of Oligonucleotides

An oligonucleotide phosphorylated at its 3' end was converted to a 3' hydrazide in 3' by treatment with an aqueous solution of hydrazine or of dihydrazide of the formula $H_2N(R1)NH_2$ at about 1 to 3 M, and at pH 4.5 at a temperature of 8°C overnight. This

19

incubation was performed in the presence of a carbodiimide type agent soluble in water such as 1-ethyl-3-(3-dimethylaminopropyl)carbodiimide at a final concentration of 0.3 M.

The derivatized oligonucleotide was then separated from the other agents and products using a standard technique for isolating oligonucleotides.

As discussed above, the mRNAs to be enriched may be treated to eliminate the 3' OH groups which may be present thereon. This may be accomplished by enzymatic ligation of sequences lacking a 3' OH, such as pCp, as described in Example 1. Alternatively, the 3' OH groups may be eliminated by alkaline hydrolysis as described in Example 8 below.

10

5

EXAMPLE 8

Elimination of 3' OH Groups of mRNA Using Alkaline Hydrolysis

In a total volume of 100 μ l of 0.1 N sodium hydroxide, 1.5 μ g mRNA is incubated for 40 to 60 minutes at 4°C. The solution is neutralized with acetic acid and precipitated with ethanol.

15

Following the optional elimination of the 3' OH groups, the diol groups at the 5' ends of the mRNAs are oxidized as described below in Example 9.

EXAMPLE 9

Oxidation of Diols of mRNA

20

Up to 1 OD unit of RNA was dissolved in 9 μ l of buffer (0.1 M sodium acetate, pH 6-7) or water and 3 μ l of freshly prepared 0.1 M sodium periodate solution. The reaction was incubated for 1 h in the dark at 4°C or room temperature. Following the incubation, the reaction was stopped by adding 4 μ l of 10% ethylene glycol. Thereafter the mixture was incubated at room temperature for 15 minutes. After ethanol precipitation, the product was resuspended in at least 10 μ l of water or appropriate buffer and dialyzed against water.

25

Following oxidation of the diol groups at the 5' ends of the mRNAs, the derivatized oligonucleotide was joined to the resulting aldehydes as described in Example 10.

20

EXAMPLE 10

Ligature of Aldehydes of mRNA to Derivatized Oligonucleotides

The oxidized mRNA was dissolved in an acidic medium such as 50 µl of sodium acetate pH 4-6. Fifty µl of a solution of the derivatized oligonucleotide were added in order to obtain an mRNA:derivatized oligonucleotide ratio of 1:20. The mixture was reduced with a borohydride and incubated for 2 h at 37°C or overnight (14 h) at 10°C. The mixture was then ethanol precipitated, resuspended in 10 µl or more of water or appropriate buffer and dialyzed against distilled water. If desired, the resulting product may be analyzed using acrylamide gel electrophoresis, HPLC analysis, or other conventional techniques.

10

15

20

25

30

5

Following the attachment of the derivatized oligonucleotide to the mRNAs, a reverse transcription reaction may be performed as described in Example 11 below.

EXAMPLE 11

Reverse Transcription of mRNAs Ligatured to Derivatized Oligonucleotides

An oligodeoxyribonucleotide was derivatized as follows. Three OD units of an oligodeoxyribonucleotide of sequence 5'ATCAAGAATTCGCACGAGACCATTA3' (SEQ ID NO:3) having 5'-OH and 3'-P ends were dissolved in 70 µl of a 1.5 M hydroxybenzotriazole solution, pH 5.3, prepared in dimethylformamide/water (75:25) containing 2 µg of 1-ethyl-3-(3-dimethylaminopropyl)carbodiimide. The mixture was incubated for 2 h 30 min at 22°C and then precipitated twice in LiClO₂/acetone. The pellet was resuspended in 200 µl of 0.25 M hydrazine and incubated at 8°C from 3 to 14 h. Following the hydrazine reaction, the mixture was precipitated twice in LiClO₂/acetone.

The messenger RNAs to be reverse transcribed were extracted from blocks of placenta having sides of 2 cm which had been stored at -80°C. The total RNA was extracted using conventional acidic phenol techniques. Oligo-dT chromatography was used to purify the mRNAs. The integrity of the mRNAs was checked by Northern-blotting.

The diol groups on 7 μ g of the placental mRNAs were oxidized as described above in Example 9. The derivatized oligonucleotide was joined to the mRNAs as described in Example 10 above except that the precipitation step was replaced by an exclusion

chromatography step to remove derivatized oligodeoxyribonucleotides which were not joined to mRNAs. Exclusion chromatography was performed as follows:

Ten ml of Ultrogel AcA34 (BioSepra#230151) gel, a mix of agarose and acrylamide, were equilibrated in 50 ml of a solution of 10 mM Tris pH 8.0, 300 mM NaCl, 1 mM EDTA, and 0.05% SDS. The mixture was allowed to sediment. The supernatant was eliminated and the gel was resuspended in 50 ml of buffer. This procedure was repeated 2 or 3 times.

5

10

15

20

25

30

A glass bead (diameter 3 mm) was introduced into a 2 ml disposable pipette (length 25 cm). The pipette was filled with the gel suspension until the height of the gel stabilized at 1 cm from the top of the pipette. The column was then equilibrated with 20 ml of equilibration buffer (10 mM Tris HCl pH 7.4, 20 mM NaCl).

Ten μ l of the mRNA which had reacted with the derivatized oligonucleotide were mixed in 39 μ l of 10 mM urea and 2 μ l of blue-glycerol buffer, which had been prepared by dissolving 5 mg of bromophenol blue in 60% glycerol (v/v), and passing the mixture through a 0.45 μ m diameter filter.

The column was then loaded with the mRNAs coupled to the oligonucleotide. As soon as the sample had penetrated, equilibration buffer was added. Hundred µl fractions were then collected. Derivatized oligonucleotide which had not been attached to mRNA appeared in fraction 16 and later fractions. Thus, fractions 3 to 15 were combined and precipitated with ethanol.

To determine whether the derivatized oligonucleotide was actually linked to mRNA, one tenth of the combined fractions were spotted twice on a nylon membrane and hybridized to a radioactive probe using conventional techniques. The ³²P labeled probe used in these hybridizations was an oligodeoxyribonucleotide of sequence 5'TAATGGTCTCGTGCGAATTCTTGAT3' (SEQ ID NO:4) anticomplementary to the derivatized oligonucleotide. A signal observed after autoradiography, indicated that the derivatized oligonucleotide had been truly joined to the mRNA.

The remaining nine tenth of the mRNAs which had reacted with the derivatized oligonucleotide was reverse transcribed as follows. A reverse transcription reaction was carried out with reverse transcriptase following the manufacturer's instructions and 50 pmol of nonamers with random sequence as primers.

5

10

15

30

To ensure that reverse transcription had been carried out through the cap structure, two types of experiments were performed.

In the first approach, after elimination of RNA of the cDNA:RNA heteroduplexes obtained from the reverse transcription reaction by an alkaline hydrolysis, a portion of the resulting single stranded cDNAs was spotted on a positively charged membrane and hybridized, using conventional methods, to a ³²P labeled probe having a sequence identical to that of the derivatized oligonucleotide. Control spots containing, 1 pmol, 100 fmol, 50 fmol, 10 fmol and 1 fmol of a control oligodeoxyribonucleotide of sequence identical to that of the derivatized oligonucleotide were included. The signal observed in the spots containing the cDNA indicated that approximately 15 fmol of the derivatized oligonucleotide had been reverse transcribed. These results demonstrate that the reverse transcription can be performed through the cap and, in particular, that reverse transcriptase crosses the 5'-P-P-P-5' bond of the cap of eukaryotic messenger RNAs.

In the second type of experiment, the single stranded cDNAs obtained from the above first strand synthesis were used as template for PCR reactions. Two types of reactions were carried out. First, specific amplification of the mRNAs for alpha globin, dehydrogenase, pp15 and elongation factor E4 were carried out using the following pairs of oligodeoxyribonucleotide primers.

20 alpha-globin

GLO-S: 5'CCG ACA AGA CCA ACG TCA AGG CCG C3' (SEQ ID NO:5)
GLO-As: 5'TCA CCA GCA GGC AGT GGC TTA GGA G 3' (SEQ ID NO:6)

dehydrogenase

3 DH-S: 5'AGT GAT TCC TGC TAC TTT GGA TGG C3' (SEQ ID NO:7)
 3 DH-As: 5'GCT TGG TCT TGT TCT GGA GTT TAG A3' (SEQ ID NO:8)

pp15

PP15-S: 5'TCC AGA ATG GGA GAC AAG CCA ATT T3' (SEQ ID NO:9)
PP15-As: 5'AGG GAG GAG GAA ACA GCG TGA GTC C3' (SEQ ID NO:10)

Elongation factor E4

EFA1-S: 5'ATG GGA AAG GAA AAG ACT CAT ATC A3' (SEQ ID NO:11) EF1A-As: 5'AGC AGC AAC AAT CAG GAC AGC ACA G3' (SEQ ID NO:12)

Second, non specific amplifications were also carried out with the antisense oligodeoxyribonucleotides of the pairs described above and with a primer derived from the sequence of the derivatized oligodeoxyribonucleotide (5'ATCAAGAATTCGCACGAGACCATTA3') (SEQ ID NO:13).

One twentieth of the following RT-PCR product samples were run on a 1.5% agarose gel and stained with ethidium bromide.

- Sample 1: The products of a PCR reaction using the globin primers of SEQ ID NOs 5 and 6 in the presence of cDNA.
- Sample 2: The products of a PCR reaction using the globin primers of SEQ ID NOs 5 and 6 in the absence of added cDNA.
- Sample 3: The products of a PCR reaction using the dehydrogenase primers of SEQ ID NOs 7 and 8 in the presence of cDNA.
 - Sample 4: The products of a PCR reaction using the dehydrogenase primers of SEQ ID NOs 7 and 8 in the absence of added cDNA.
- Sample 5: The products of a PCR reaction using the pp15 primers of SEQ ID NOs 9 and 10 in the presence of cDNA.
 - Sample 6: The products of a PCR reaction using the pp15 primers of SEQ ID NOs 9 and 10 in the absence of added cDNA.
 - Sample 7: The products of a PCR reaction using the EIF4 primers of SEQ ID NOs 11 and 12 in the presence of added cDNA.
- Sample 8: The products of a PCR reaction using the EIF4 primers of SEQ ID NOs 11 and 12 in the absence of added cDNA.

A band of the size expected for the PCR product was observed only in samples 1, 3, 5 and 7, thus indicating the presence of the corresponding sequence in the cDNA population.

PCR reactions were also carried out with the antisense oligonucleotides of the globin and dehydrogenase primers (SEQ ID NOs 6 and 8) and an oligonucleotide whose sequence corresponds to that of the derivatized oligonucleotide. The presence of PCR products of the expected size in the samples equivalent to above samples 1 and 3 indicated that the derivatized oligonucleotide had been linked to mRNA.

5

10

15

20

25

30

The above examples summarize the chemical procedure for enriching mRNAs for those having intact 5' ends as illustrated in Figure 1. Further detail regarding the chemical approaches for obtaining such mRNAs are disclosed in International Application No. WO96/34981, published November 7, 1996, which is incorporated herein by reference. Strategies based on the above chemical modifications to the 5' cap structure may be utilized to generate cDNAs selected to include the 5' ends of the mRNAs from which they derived. In one version of such procedures, the 5' ends of the mRNAs are modified as described Thereafter, a reverse transcription reaction is conducted to extend a primer complementary to the 5' end of the mRNA. Single stranded RNAs are eliminated to obtain a population of cDNA/mRNA heteroduplexes in which the mRNA includes an intact 5' end. The resulting heteroduplexes may be captured on a solid phase coated with a molecule capable of interacting with the molecule used to derivatize the 5' end of the mRNA. Thereafter, the strands of the heteroduplexes are separated to recover single stranded first cDNA strands which include the 5' end of the mRNA. Second strand cDNA synthesis may then proceed using conventional techniques. For example, the procedures disclosed in WO 96/34981 or in Carninci. et al., Genomics 37:327-336, 1996, the disclosures of which are incorporated herein by reference, may be employed to select cDNAs which include the sequence derived from the 5' end of the coding sequence of the mRNA.

Following ligation of the oligonucleotide tag to the 5' cap of the mRNA, a reverse transcription reaction is conducted to extend a primer complementary to the mRNA to the 5' end of the mRNA. Following elimination of the RNA component of the resulting heteroduplex using standard techniques, second strand cDNA synthesis is conducted with a primer complementary to the oligonucleotide tag.

2. Enzymatic Methods for Obtaining mRNAs having Intact 5' Ends

Other techniques for selecting cDNAs extending to the 5' end of the mRNA from which they are derived are fully enzymatic. Some versions of these techniques are disclosed in Dumas Milne Edwards J.B. (Doctoral Thesis of Paris VI University, Le clonage des ADNc

complets: difficultes et perspectives nouvelles. Apports pour l'etude de la regulation de l'expression de la tryptophane hydroxylase de rat, 20 Dec. 1993), EPO 625572 and Kato et al., Gene 150:243-250, 1994, the disclosures of which are incorporated herein by reference.

Briefly, in such approaches, isolated mRNA is treated with alkaline phosphatase to remove the phosphate groups present on the 5' ends of uncapped incomplete mRNAs. Following this procedure, the cap present on full length mRNAs is enzymatically removed with a decapping enzyme such as T4 polynucleotide kinase or tobacco acid pyrophosphatase. An oligonucleotide, which may be either a DNA oligonucleotide or a DNA-RNA hybrid oligonucleotide having RNA at its 3' end, is then ligated to the phosphate present at the 5' end of the decapped mRNA using T4 RNA ligase. The oligonucleotide may include a restriction site to facilitate cloning of the cDNAs following their synthesis. Example 12 below describes one enzymatic method based on the doctoral thesis of Dumas.

EXAMPLE 12

15

20

25

30

10

5

Enzymatic Approach for Obtaining 5' ESTs

Twenty micrograms of PolyA+ RNA were dephosphorylated using Calf Intestinal Phosphatase (Biolabs). After a phenol chloroform extraction, the cap structure of mRNA was hydrolysed using the Tobacco Acid Pyrophosphatase (purified as described by Shinshi et al., Biochemistry 15: 2185-2190, 1976) and a hemi 5'DNA/RNA-3' oligonucleotide having an unphosphorylated 5' end, a stretch of adenosine ribophosphate at the 3' end, and an EcoRI site near the 5' end was ligated to the 5'P ends of mRNA using the T4 RNA ligase (Biolabs). Oligonucleotides suitable for use in this procedure are preferably 30 to 50 bases in length. Oligonucleotides having an unphosphorylated 5' end may be synthesized by adding a fluorochrome at the 5' end. The inclusion of a stretch of adenosine ribophosphates at the 3' end of the oligonucleotide increases ligation efficiency. It will be appreciated that the oligonucleotide may contain cloning sites other than EcoRI.

Following ligation of the oligonucleotide to the phosphate present at the 5' end of the decapped mRNA, first and second strand cDNA synthesis is carried out using conventional methods or those specified in EP0 625,572 and Kato et al. supra, and Dumas Milne Edwards, supra, the disclosures of which are incorporated herein by

26

reference. The resulting cDNA may then be ligated into vectors such as those disclosed in Kato et al., supra or other nucleic acid vectors known to those skilled in the art using techniques such as those described in Sambrook et al., Molecular Cloning: A Laboratory Manual 2d Ed., Cold Spring Harbor Laboratory Press, 1989, the disclosure of which is incorporated herein by reference.

II. Obtention and Characterization of the 5' ESTs of the Present Invention

The 5' ESTs of the present invention were obtained using the aforementioned chemical and enzymatic approaches for enriching mRNAs for those having intact 5' ends as decribed below.

1. Obtention of 5' ESTS Using mRNAs with Intact 5' Ends

First, mRNAs were prepared as described in Example 13 below.

15

20

25

30

10

5

EXAMPLE 13

Preparation of mRNA With Intact 5' Ends

Total human RNAs or polyA RNAs derived from 29 different tissues were respectively purchased from LABIMO and CLONTECH and used to generate 44 cDNA libraries as follows. The purchased RNA had been isolated from cells or tissues using acid guanidium thiocyanate-phenol-chloroform extraction (Chomczyniski and Sacchi, *Analytical Biochemistry* 162:156-159, 1987). PolyA RNA was isolated from total RNA (LABIMO) by two passes of oligo dT chromatography, as described by Aviv and Leder, *Proc. Natl. Acad. Sci. USA* 69:1408-1412, 1972 in order to eliminate ribosomal RNA.

The quality and the integrity of the polyA+ RNAs were checked. Northern blots hybridized with a globin probe were used to confirm that the mRNAs were not degraded. Contamination of the polyA⁺ mRNAs by ribosomal sequences was checked using Northern blots and a probe derived from the sequence of the 28S rRNA. Preparations of mRNAs with less than 5% of rRNAs were used in library construction. To avoid constructing libraries with RNAs contaminated by exogenous sequences (prokaryotic or fungal), the presence of bacterial 16S ribosomal sequences or of two highly expressed fungal mRNAs was examined using PCR.

27

Following preparation of the mRNAs, the above described chemical and/or the enzymatic procedures for enriching mRNAs for thoses having intact 5' ends were employed to obtain 5' ESTs from various tissues. In both approaches, an oligonucleotide tag was attached to the 5' ends of the mRNAs. The oligonucleotide tag had an EcoRI site therein to facilitate later cloning procedures. To facilitate the processing of single stranded and double stranded cDNA obtained in the construction of the librairies, the same nucleotidic sequence was used to design the ligated oligonucleotide in both chemical and enzymatic approaches. Nevertheless, in the chemical procedure, the tag used was an oligodeoxyribonucleotide which was linked to the cap of the mRNA whereas in the enzymatic ligation, the tag was a chimeric hemi 5'DNA/RNA3' oligonucleotide which was ligated to the 5' end of decapped mRNA as described in example 12.

5

10

15

20

25

30

Following attachment of the oligonucleotide tag to the mRNA by either the chemical or enzymatic methods, the integrity of the mRNA was examined by performing a Northern blot with 200 to 500 ng of mRNA using a probe complementary to the oligonucleotide tag before performing the first strand synthesis as described in example 14.

EXAMPLE 14

cDNA Synthesis Using mRNA Templates Having Intact 5' Ends

For the mRNAs joined to oligonucleotide tags using both the chemical and enzymatic methods, first strand cDNA synthesis was performed using the Superscript II (Gibco BRL) or the Rnase H Minus M-MLV (Promega) reverse transcriptase with random nonamers as primers. In order to protect internal EcoRI sites in the cDNA from digestion at later steps in the procedure, methylated dCTP was used for first strand synthesis. After removal of RNA by an alkaline hydrolysis, the first strand of cDNA was precipitated using isopropanol in order to eliminate residual primers.

For both the chemical and the enzymatic methods, the second strand of the cDNA was synthesized with a Klenow fragment using a primer corresponding to the 5' end of the ligated oligonucleotide described in Example 12. Preferably, the primer is 20-25 bases in length. Methylated dCTP was also used for second strand synthesis in order to protect internal EcoRI sites in the cDNA from digestion during the cloning process.

5

10

15

Following cDNA synthesis, the cDNAs were cloned into pBlueScript as described in Example 15 below.

EXAMPLE 15

Cloning of cDNAsderived from mRNA with intact 5' ends into BlueScript

Following second strand synthesis, the ends of the cDNA were blunted with T4 DNA polymerase (Biolabs) and the cDNA was digested with EcoRI. Since methylated dCTP was used during cDNA synthesis, the EcoRI site present in the tag was the only hemi-methylated site, hence the only site susceptible to EcoRI digestion. The cDNA was then size fractionated using exclusion chromatography (AcA, Biosepra) and fractions corresponding to cDNAs of more than 150 bp were pooled and ethanol precipitated. The cDNA was directionally cloned into the SmaI and EcoRI ends of the phagemid pBlueScript vector (Stratagene). The ligation mixture was electroporated into bacteria and propagated under appropriate antibiotic selection.

Clones containing the oligonucleotide tag attached were then selected as described in Example 16 below.

EXAMPLE 16

Selection of Clones Having the Oligonucleotide Tag Attached Thereto

The plasmid DNAs containing 5' EST libraries made as described above were purified (Qiagen). A positive selection of the tagged clones was performed as follows. Briefly, in this selection procedure, the plasmid DNA was converted to single stranded DNA using gene II endonuclease of the phage F1 in combination with an exonuclease (Chang et al., Gene 127:95-8, 1993) such as exonuclease III or T7 gene 6 exonuclease. The resulting single stranded DNA was then purified using paramagnetic beads as described by Fry et al., Biotechniques, 13: 124-131, 1992. In this procedure, the single stranded DNA was hybridized with a biotinylated oligonucleotide having a sequence corresponding to the 3' end of the oligonucleotide described in Example 13. Preferably, the primer has a length of 20-25 bases. Clones including a sequence complementary to the biotinylated oligonucleotide were captured by incubation with streptavidin coated magnetic beads followed by magnetic selection. After capture of the positive clones, the plasmid DNA was released from the

29

magnetic beads and converted into double stranded DNA using a DNA polymerase such as the ThermoSequenase obtained from Amersham Pharmacia Biotech. Alternatively, protocoles such as the one described in the Gene Trapper kit available from Gibco BRL may be used. The double stranded DNA was then electroporated into bacteria. The percentage of positive clones having the 5' tag oligonucleotide was estimated to typically rank between 90 and 98% using dot blot analysis.

Following electroporation, the libraries were ordered in 384-microtiter plates (MTP). A copy of the MTP was stored for future needs. Then the libraries were transferred into 96 MTP and sequenced as described below.

10

15

20

25

5

EXAMPLE 17

Sequencing of Inserts in Selected Clones

Plasmid inserts were first amplified by PCR on PE 9600 thermocyclers (Perkin-Elmer, Applied Biosystems Division, Foster City, CA), using standard SETA-A and SETA-B primers (Genset SA), AmpliTaqGold (Perkin-Elmer), dNTPs (Boehringer), buffer and cycling conditions as recommended by the Perkin-Elmer Corporation.

PCR products were then sequenced using automatic ABI Prism 377 sequencers (Perkin Elmer). Sequencing reactions were performed using PE 9600 thermocyclers with standard dye-primer chemistry and ThermoSequenase (Amersham Pharmacia Biotech). The primers used were either T7 or 21M13 (available from Genset SA) as appropriate. The primers were labeled with the JOE, FAM, ROX and TAMRA dyes. The dNTPs and ddNTPs used in the sequencing reactions were purchased from Boehringer. Sequencing buffer, reagent concentrations and cycling conditions were as recommended by Amersham.

Following the sequencing reaction, the samples were precipitated with ethanol, resuspended in formamide loading buffer, and loaded on a standard 4% acrylamide gel. Electrophoresis was performed for 2.5 hours at 3000V on an ABI 377 sequencer, and the sequence data were collected and analyzed using the ABI Prism DNA Sequencing Analysis Software, version 2.1.2.

30

2. Computer analysis of the Obtained 5' ESTs: Construction of NetGene and SignalTag databases

The sequence data from the 44 cDNA libraries made as described above were transferred to a proprietary database, where quality control and validation steps were performed. A proprietary base-caller, working using a Unix system, automatically flagged suspect peaks, taking into account the shape of the peaks, the inter-peak resolution, and the noise level. The proprietary base-caller also performed an automatic trimming. Any stretch of 25 or fewer bases having more than 4 suspect peaks was considered unreliable and was discarded. Sequences corresponding to cloning vector or ligation oligonucleotides were automatically removed from the EST sequences. However, the resulting EST sequences may contain 1 to 5 bases belonging to the above mentioned sequences at their 5' end. If needed, these can easily be removed on a case to case basis.

Following sequencing as described above, the sequences of the 5' ESTs were entered in NetGeneTM, a proprietary database called for storage and manipulation as described below. It will be appreciated by those skilled in the art that the data could be stored and manipulated on any medium which can be read and accessed by a computer. Computer readable media include magnetically, optically, or electronically readable media. For example, the computer readable media may be a hard disc, a floppy disc, a magnetic tape, CD-ROM, RAM, or ROM as well as other types of other media known to those skilled in the art.

20

5

10

15

In addition, the sequence data may be stored and manipulated in a variety of data processor programs in a diversity of formats. For instance, the sequence data may be stored as text in a word processing file, such as Microsoft WORD or WORDPERFECT or as an ASCII file in a variety of database programs familiar to those of skill in the art, such as DB2, SYBASE, or ORACLE.

25

30

The computer readable media on which the sequence information is stored may be in a personal computer, a network, a server or other computer systems known to those skilled in the art. The computer or other system preferably includes the storage media described above, and a processor for accessing and manipulating the sequence data. Once the sequence data has been stored, it may be manipulated and searched to locate those stored sequences which contain a desired nucleic acid sequence or which encode a protein having a particular functional domain. For example, the stored sequence information may be compared to other

known sequences to identify homologies, motifs implicated in biological function, or structural motifs.

Programs which may be used to search or compare the stored sequences include the MacPattern (EMBL), BLAST, and BLAST2 program series (NCBI), basic local alignment search tool programs for nucleotide (BLASTN) and peptide (BLASTX) comparisons (Altschul et al, J. Mol. Biol. 215: 403, 1990) and FASTA (Pearson and Lipman, Proc. Natl. Acad. Sci. USA 85: 2444, 1988). The BLAST programs then extend the alignments on the basis of defined match and mismatch criteria.

5

10

15

Motifs which may be detected using the above programs and those described in Example 28 include sequences encoding leucine zippers, helix-turn-helix motifs, glycosylation sites, ubiquitination sites, alpha helices, and beta sheets, signal sequences encoding signal peptides which direct the secretion of the encoded proteins, sequences implicated in transcription regulation such as homeoboxes, acidic stretches, enzymatic active sites, substrate binding sites, and enzymatic cleavage sites.

Before searching the cDNAs in the NetGene™ database for sequence motifs of interest, cDNAs derived from mRNAs which were not of interest were identified and eliminated from further consideration as described in Example 18 below.

EXAMPLE 18

20 <u>Elimination of Undesired Sequences from Further Consideration</u>

5' ESTs in the NetGene™ database which were derived from undesired sequences such as transfer RNAs, ribosomal RNAs, mitochondrial RNAs, prokaryotic RNAs, fungal RNAs, Alu sequences, L1 sequences, or repeat sequences were identified using the FASTA and BLASTN programs with the parameters listed in Table I.

To eliminate 5' ESTs encoding tRNAs from further consideration, the 5' EST sequences were compared to the sequences of 1190 known tRNAs obtained from EMBL release 38, of which 100 were human. The comparison was performed using FASTA on both strands of the 5' ESTs. Sequences having more than 80% homology over more than 60 nucleotides were identified as tRNA. Of the 144,341 sequences screened, 26 were identified as tRNAs and eliminated from further consideration.

32

To eliminate 5' ESTs encoding rRNAs from further consideration, the 5' EST sequences were compared to the sequences of 2497 known rRNAs obtained from EMBL release 38, of which 73 were human. The comparison was performed using BLASTN on both strands of the 5' ESTs with the parameter S=108. Sequences having more than 80% homology over stretches longer than 40 nucleotides were identified as rRNAs. Of the 144,341 sequences screened, 3,312 were identified as rRNAs and eliminated from further consideration.

5

10

15

20

25

30

To eliminate 5' ESTs encoding mtRNAs from further consideration, the 5' EST sequences were compared to the sequences of the two known mitochondrial genomes for which the entire genomic sequences are available and all sequences transcribed from these mitochondrial genomes including tRNAs, rRNAs, and mRNAs for a total of 38 sequences. The comparison was performed using BLASTN on both strands of the 5' ESTs with the parameter S=108. Sequences having more than 80% homology over stretches longer than 40 nucleotides were identified as mtRNAs. Of the 144,341 sequences screened, 6,110 were identified as mtRNAs and eliminated from further consideration.

Sequences which might have resulted from exogenous contaminants were eliminated from further consideration by comparing the 5' EST sequences to release 46 of the EMBL bacterial and fungal divisions using BLASTN with the parameter S=144. All sequences having more than 90% homology over at least 40 nucleotides were identified as exogenous contaminants. Of the 42 cDNA libraries examined, the average percentages of prokaryotic and fungal sequences contained therein were 0.2% and 0.5% respectively. Among these sequences, only one could be identified as a sequence specific to fungi. The others were either fungal or prokaryotic sequences having homologies with vertebrate sequences or including repeat sequences which had not been masked during the electronic comparison.

In addition, the 5' ESTs were compared to 6093 Alu sequences and 1115 L1 sequences to mask 5' ESTs containing such repeat sequences. 5' ESTs including THE and MER repeats, SSTR sequences or satellite, micro-satellite, or telomeric repeats were also eliminated from further consideration. On average, 11.5% of the sequences in the libraries contained repeat sequences. Of this 11.5%, 7% contained Alu repeats, 3.3% contained L1 repeats and the remaining 1.2% were derived from the other screened types of repetitive sequences. These percentages are consistent with those found in cDNA libraries prepared by

other groups. For example, the cDNA libraries of Adams *et al.* contained between 0% and 7.4% Alu repeats depending on the source of the RNA which was used to prepare the cDNA library (Adams *et al.*, *Nature* 377:174, 1996).

The sequences of those 5' ESTs remaining after the elimination of undesirable sequences were compared with the sequences of known human mRNAs to determine the accuracy of the sequencing procedures described above.

EXAMPLE 19

Measurement of Sequencing Accuracy by Comparison to Known Sequences

10

15

20

25

To further determine the accuracy of the sequencing procedure described above, the sequences of 5' ESTs derived from known sequences were identified and compared to the original known sequences. First, a FASTA analysis with overhangs shorter than 5 bp on both ends was conducted on the 5' ESTs to identify those matching an entry in the public human mRNA database. The 6655 5' ESTs which matched a known human mRNA were then realigned with their cognate mRNA and dynamic programming was used to include substitutions, insertions, and deletions in the list of "errors" which would be recognized. Errors occurring in the last 10 bases of the 5' EST sequences were ignored to avoid the inclusion of spurious cloning sites in the analysis of sequencing accuracy.

This analysis revealed that the sequences incorporated in the NetGene™ database had an accuracy of more than 99.5%.

To determine the efficiency with which the above selection procedures select cDNAs which include the 5' ends of their corresponding mRNAs, the following analysis was performed.

EXAMPLE 20

Determination of Efficiency of 5' EST Selection

To determine the efficiency at which the above selection procedures isolated 5' ESTs which included sequences close to the 5' end of the mRNAs from which they derived, the sequences of the ends of the 5' ESTs derived from the elongation factor 1 subunit α and

WO 99/06439

5

10

15

20

25

30

ferritin heavy chain genes were compared to the known cDNA sequences of these genes. Since the transcription start sites of both genes are well characterized, they may be used to determine the percentage of derived 5' ESTs which included the authentic transcription start sites.

For both genes, more than 95% of the obtained 5' ESTs actually included sequences close to or upstream of the 5' end of the corresponding mRNAs.

To extend the analysis of the reliability of the procedures for isolating 5' ESTs from ESTs in the NetGeneTM database, a similar analysis was conducted using a database composed of human mRNA sequences extracted from GenBank database release 97 for comparison. The 5' ends of more than 85% of 5' ESTs derived from mRNAs included in the GeneBank database were located close to the 5' ends of the known sequence. As some of the mRNA sequences available in the GenBank database are deduced from genomic sequences, a 5' end matching with these sequences will be counted as an internal match. Thus, the method used here underestimates the yield of ESTs including the authentic 5' ends of their corresponding mRNAs.

The EST libraries made above included multiple 5' ESTs derived from the same mRNA. The sequences of such 5' ESTs were compared to one another and the longest 5' ESTs for each mRNA were identified. Overlapping cDNAs were assembled into continuous sequences (contigs). The resulting continuous sequences were then compared to public databases to gauge their similarity to known sequences, as described in Example 21 below.

EXAMPLE 21

Clustering of the 5' ESTs and Calculation of Novelty Indices for cDNA Libraries

For each sequenced EST library, the sequences were clustered by the 5' end. Each sequence in the library was compared to the others with BLASTN2 (direct strand, parameters S=107). ESTs with High Scoring Segment Pairs (HSPs) at least 25 bp long, having 95% identical bases and beginning closer than 10 bp from each EST 5' end were grouped. The longest sequence found in the cluster was used as representative of the group. A global clustering between libraries was then performed leading to the definition of super-contigs.

35

To assess the yield of new sequences within the EST libraries, a novelty rate (NR) was defined as: NR= 100 X (Number of new unique sequences found in the library/Total number of sequences from the library). Typically, novelty rating ranged between 10% and 41% depending on the tissue from which the EST library was obtained. For most of the libraries, the random sequencing of 5' EST libraries was pursued until the novelty rate reached 20%.

Following characterization as described above, the collection of 5' ESTs in NetGeneTM was screened to identify those 5' ESTs bearing potential signal sequences as described in Example 22 below.

EXAMPLE 22

Identification of Potential Signal Sequences in 5' ESTs

The 5' ESTs in the NetGeneTM database were screened to identify those having an uninterrupted open reading frame (ORF) longer than 45 nucleotides beginning with an ATG codon and extending to the end of the EST. Approximately half of the cDNA sequences in NetGeneTM contained such an ORF. The ORFs of these 5' ESTs were then searched to identify potential signal motifs using slight modifications of the procedures disclosed in Von Heijne, *Nucleic Acids Res.* 14:4683-4690, 1986, the disclosure of which is incorporated herein by reference. Those 5' EST sequences encoding a stretch of at least 15 amino acid long with a score of at least 3.5 in the Von Heijne signal peptide identification matrix were considered to possess a signal sequence. Those 5' ESTs which matched a known human mRNA or EST sequence and had a 5' end more than 20 nucleotides downstream of the known 5' end were excluded from further analysis. The remaining cDNAs having signal sequences therein were included in a database called SignalTagTM.

To confirm the accuracy of the above method for identifying signal sequences, the analysis of Example 23 was performed.

5

10

15

20

25

EXAMPLE 23

Confirmation of Accuracy of Identification of Potential Signal Sequences in 5' ESTs

The accuracy of the above procedure for identifying signal sequences encoding signal peptides was evaluated by applying the method to the 43 amino acids located at the N terminus of all human SwissProt proteins. The computed Von Heijne score for each protein was compared with the known characterization of the protein as being a secreted protein or a non-secreted protein. In this manner, the number of non-secreted proteins having a score higher than 3.5 (false positives) and the number of secreted proteins having a score lower than 3.5 (false negatives) could be calculated.

Using the results of the above analysis, the probability that a peptide encoded by the 5' region of the mRNA is in fact a genuine signal peptide based on its Von Heijne's score was calculated based on either the assumption that 10% of human proteins are secreted or the assumption that 20% of human proteins are secreted. The results of this analysis are shown in Figure 2 and table IV.

Using the above method of identification of secretory proteins, 5' ESTs of the following polypeptides known to be secreted were obtained: human glucagon, gamma interferon induced monokine precursor, secreted cyclophilin-like protein, human pleiotropin, and human biotinidase precursor. Thus, the above method successfully identified those 5' ESTs which encode a signal peptide.

20

25

30

15

5

10

To confirm that the signal peptide encoded by the 5' ESTs actually functions as a signal peptide, the signal sequences from the 5' ESTs may be cloned into a vector designed for the identification of signal peptides. Such vectors are designed to confer the ability to grow in selective medium only to host cells containing a vector with an operably linked signal sequence. For example, to confirm that a 5' EST encodes a genuine signal peptide, the signal sequence of the 5' EST may be inserted upstream and in frame with a non-secreted form of the yeast invertase gene in signal peptide selection vectors such as those described in U.S. Patent No. 5,536,637, the disclosure of which is incorporated herein by reference. Growth of host cells containing signal sequence selection vectors with the correctly inserted 5' EST signal sequence confirms that the 5' EST encodes a genuine signal peptide.

37

Alternatively, the presence of a signal peptide may be confirmed by cloning the extended cDNAs obtained using the ESTs into expression vectors such as pXT1 (as described below in example 30), or by constructing promoter-signal sequence-reporter gene vectors which encode fusion proteins between the signal peptide and an assayable reporter protein. After introduction of these vectors into a suitable host cell, such as COS cells or NIH 3T3 cells, the growth medium may be harvested and analyzed for the presence of the secreted protein. The medium from these cells is compared to the medium from control cells containing vectors lacking the signal sequence or extended cDNA insert to identify vectors which encode a functional signal peptide or an authentic secreted protein.

Those 5' ESTs which encoded a signal peptide, as determined by the method of Example 22 above, were further grouped into four categories based on their homology to known sequences as described in Example 24 below.

EXAMPLE 24

15

25

30

10

5

Categorization of 5' ESTs Encoding a Signal Peptide

Those 5' ESTs having a sequence not matching any known vertebrate sequence nor any publicly available EST sequence were designated "new." Of the sequences in the SignalTag[™] database, 947 of the 5' ESTs having a Von Heijne's score of at least 3.5 fell into this category.

Those 5' ESTs having a sequence not matching any vertebrate sequence but matching a publicly known EST were designated "EST-ext", provided that the known EST sequence was extended by at least 40 nucleotides in the 5' direction. Of the sequences in the SignalTag[™] database, 150 of the 5' ESTs having a Von Heijne's score of at least 3.5 fell into this category.

Those ESTs not matching any vertebrate sequence but matching a publicly known EST without extending the known EST by at least 40 nucleotides in the 5' direction were designated "EST." Of the sequences in the SignalTag[™] database, 599 of the 5' ESTs having a Von Heijne's score of at least 3.5 fell into this category.

Those 5' ESTs matching a human mRNA sequence but extending the known sequence by at least 40 nucleotides in the 5' direction were designated "VERT-ext." Of the sequences in the SignalTag[™] database, 23 of the 5' ESTs having a Von Heijne's score of at

least 3.5 fell into this category. Included in this category was a 5' EST which extended the known sequence of the human translocase mRNA by more than 200 bases in the 5' direction. A 5' EST which extended the sequence of a human tumor suppressor gene in the 5' direction was also identified.

Table V shows the distribution of 5' ESTs in each category and the number of 5' ESTs in each category having a given minimum von Heijne's score.

3. Evaluation of Spatial and Temporal Expression of mRNAs Corresponding to the 5'ESTs or Extended cDNAs

10

5

Each of the 5' ESTs was also categorized based on the tissue from which its corresponding mRNA was obtained, as described below in Example 25.

EXAMPLE 25

15

20

Categorization of Expression Patterns

Table VI shows the distribution of 5' ESTs in each of the above defined category with respect to the tissue from which the 5'ESTs of the corresponding mRNA were obtained.

Table II provides the sequence identification numbers of 5' EST sequences derived from endoderm, the categories in which these sequences fall, and the von Heijne's score of the signal peptides which they encode. The 5' EST sequences and the amino acid sequences they encode are provided in the appended sequence listings. Table III provides the sequence ID numbers of the 5' ESTs and the sequences of the signal peptides which they encode. The sequences of the 5' ESTs and the polypeptides they encode are provided in the sequence listing appended hereto.

25

30

The sequences of DNA SEQ ID NOs: 38-184 can readily be screened for any errors therein and any sequence ambiguities can be resolved by resequencing a fragment containing such errors or ambiguities on both strands. Such fragments may be obtained from the plasmids stored in the inventors' laboratory or can be isolated using the techniques described herein. Resolution of any such ambiguities or errors may be facilitated by using primers which hybridize to sequences located close to the ambiguous or erroneous sequences. For example, the primers may hybridize to sequences within 50-75 bases of the ambiguity or

39

error. Upon resolution of an error or ambiguity, the corresponding corrections can be made in the protein sequences encoded by the DNA containing the error or ambiguity.

In addition to categorizing the 5' ESTs with respect to their tissue of origin, the spatial and temporal expression patterns of the mRNAs corresponding to the 5' ESTs, as well as their expression levels, may be determined as described in Example 26 below. Characterization of the spatial and temporal expression patterns and expression levels of these mRNAs is useful for constructing expression vectors capable of producing a desired level of gene product in a desired spatial or temporal manner, as will be discussed in more detail below.

5

10

15

20

25

30

Furthermore, 5' ESTs whose corresponding mRNAs are associated with disease states may also be identified. For example, a particular disease may result from the lack of expression, over expression, or under expression of an mRNA corresponding to a 5' EST. By comparing mRNA expression patterns and quantities in samples taken from healthy individuals with those from individuals suffering from a particular disease, 5' ESTs responsible for the disease may be identified.

It will be appreciated that the results of the above characterization procedures for 5' ESTs also apply to extended cDNAs (obtainable as described below) which contain sequences adjacent to the 5' ESTs. It will also be appreciated that if desired, characterization may be delayed until extended cDNAs have been obtained rather than characterizing the ESTs themselves.

EXAMPLE 26

Evaluation of Expression Levels and Patterns of mRNAs

Corresponding to 5' ESTs or Extended cDNAs

Expression levels and patterns of mRNAs corresponding to 5' ESTs or extended cDNAs (obtainable as described below in example 27) may be analyzed by solution hybridization with long probes as described in International Patent Application No. WO 97/05277, the entire contents of which are hereby incorporated by reference. Briefly, a 5' EST, extended cDNA, or fragment thereof corresponding to the gene encoding the mRNA to be characterized is inserted at a cloning site immediately downstream of a bacteriophage (T3,

T7 or SP6) RNA polymerase promoter to produce antisense RNA. Preferably, the 5' EST or extended cDNA has 100 or more nucleotides. The plasmid is linearized and transcribed in the presence of ribonucleotides comprising modified ribonucleotides (*i.e.* biotin-UTP and DIG-UTP). An excess of this doubly labeled RNA is hybridized in solution with mRNA isolated from cells or tissues of interest. The hybridizations are performed under standard stringent conditions (40-50°C for 16 hours in an 80% formamide, 0.4 M NaCl buffer, pH 7-8). The unhybridized probe is removed by digestion with ribonucleases specific for single-stranded RNA (*i.e.* RNases CL3, T1, Phy M, U2 or A). The presence of the biotin-UTP modification enables capture of the hybrid on a microtitration plate coated with streptavidin. The presence of the DIG modification enables the hybrid to be detected and quantified by ELISA using an anti-DIG antibody coupled to alkaline phosphatase.

5

10

15

20

The 5' ESTs, extended cDNAs, or fragments thereof may also be tagged with nucleotide sequences for the serial analysis of gene expression (SAGE) as disclosed in UK Patent Application No. 2 305 241 A, the entire contents of which are incorporated by reference. In this method, cDNAs are prepared from a cell, tissue, organism or other source of nucleic acid for which gene expression patterns must be determined. The resulting cDNAs are separated into two pools. The cDNAs in each pool are cleaved with a first restriction endonuclease, called an anchoring enzyme, having a recognition site which is likely to be present at least once in most cDNAs. The fragments which contain the 5' or 3' most region of the cleaved cDNA are isolated by binding to a capture medium such as streptavidin coated beads. A first oligonucleotide linker having a first sequence for hybridization of an amplification primer and an internal restriction site for a so-called tagging endonuclease is ligated to the digested cDNAs in the first pool. Digestion with the second endonuclease produces short tag fragments from the cDNAs.

A second oligonucleotide having a second sequence for hybridization of an amplification primer and an internal restriction site is ligated to the digested cDNAs in the second pool. The cDNA fragments in the second pool are also digested with the tagging endonuclease to generate short tag fragments derived from the cDNAs in the second pool. The tags resulting from digestion of the first and second pools with the anchoring enzyme and the tagging endonuclease are ligated to one another to produce so-called ditags. In some embodiments, the ditags are concatamerized to produce ligation products containing from 2

41

to 200 ditags. The tag sequences are then determined and compared to the sequences of the 5' ESTs or extended cDNAs to determine which 5' ESTs or extended cDNAs are expressed in the cell, tissue, organism, or other source of nucleic acids from which the tags were derived. In this way, the expression pattern of the 5' ESTs or extended cDNAs in the cell, tissue, organism, or other source of nucleic acids is obtained.

5

10

15

20

25

30

Quantitative analysis of gene expression may also be performed using arrays. As used herein, the term array means a one dimensional, two dimensional, or multidimensional arrangement of full length cDNAs (*i.e.* extended cDNAs which include the coding sequence for the signal peptide, the coding sequence for the mature protein, and a stop codon), extended cDNAs, 5' ESTs or fragments thereof of sufficient length to permit specific detection of gene expression. Preferably, the fragments are at least 15 nucleotides in length. More preferably, the fragments are at least 100 nucleotide long. More preferably, the fragments are more than 100 nucleotides in length. In some embodiments, the fragments may be more than 500 nucleotide long.

For example, quantitative analysis of gene expression may be performed with full length cDNAs as defined below, extended cDNAs, 5' ESTs, or fragments thereof in a complementary DNA microarray as described by Schena et al. (Science 270:467-470, 1995; Proc. Natl. Acad. Sci. U.S.A. 93:10614-10619, 1996). Full length cDNAs, extended cDNAs, 5' ESTs or fragments thereof are amplified by PCR and arrayed from 96-well microtiter plates onto silylated microscope slides using high-speed robotics. Printed arrays are incubated in a humid chamber to allow rehydration of the array elements and rinsed, once in 0.2% SDS for 1 min, twice in water for 1 min and once for 5 min in sodium borohydride solution. The arrays are submerged in water for 2 min at 95°C, transferred into 0.2% SDS for 1 min, rinsed twice with water, air dried and stored in the dark at 25°C.

Cell or tissue mRNA is isolated or commercially obtained and probes are prepared by a single round of reverse transcription. Probes are hybridized to 1 cm² microarrays under a 14 x 14 mm glass coverslip for 6-12 hours at 60°C. Arrays are washed for 5 min at 25°C in low stringency wash buffer (1 x SSC/0.2% SDS), then for 10 min at room temperature in high stringency wash buffer (0.1 x SSC/0.2% SDS). Arrays are scanned in 0.1 x SSC using a fluorescence laser scanning device fitted with a custom filter set. Accurate differential

expression measurements are obtained by taking the average of the ratios of two independent hybridizations.

Quantitative analysis of the expression of genes may also be performed with full length cDNAs, extended cDNAs, 5' ESTs, or fragments thereof in complementary DNA arrays as described by Pietu et al.. (Genome Research 6:492-503, 1996). The full length cDNAs, extended cDNAs, 5' ESTs or fragments thereof are PCR amplified and spotted on membranes. Then, mRNAs originating from various tissues or cells are labeled with radioactive nucleotides. After hybridization and washing in controlled conditions, the hybridized mRNAs are detected by phospho-imaging or autoradiography. Duplicate experiments are performed and a quantitative analysis of differentially expressed mRNAs is then performed.

Alternatively, expression analysis of the 5' ESTs or extended cDNAs can be done through high density nucleotide arrays as described by Lockhart et al. (Nature Biotechnology 14: 1675-1680, 1996) and Sosnowsky et al. (Proc. Natl. Acad. Sci. 94:1119-1123, 1997). Oligonucleotides of 15-50 nucleotides corresponding to sequences of the 5' ESTs or extended cDNAs are synthesized directly on the chip (Lockhart et al., supra) or synthesized and then addressed to the chip (Sosnowsky et al., supra). Preferably, the oligonucleotides are about 20 nucleotides in length.

cDNA probes labeled with an appropriate compound, such as biotin, digoxigenin or fluorescent dye, are synthesized from the appropriate mRNA population and then randomly fragmented to an average size of 50 to 100 nucleotides. The said probes are then hybridized to the chip. After washing as described in Lockhart et al, supra and application of different electric fields (Sonowsky et al, supra.), the dyes or labeling compounds are detected and quantified. Duplicate hybridizations are performed. Comparative analysis of the intensity of the signal originating from cDNA probes on the same target oligonucleotide in different cDNA samples indicates a differential expression of the mRNA corresponding to the 5' EST or extended cDNA from which the oligonucleotide sequence has been designed.

5

10

15

20

25

III. Use of 5' ESTs to Clone Extended cDNAs and to Clone the Corresponding Genomic DNAs

Once 5' ESTs which include the 5' end of the corresponding mRNAs have been selected using the procedures described above, they can be utilized to isolate extended cDNAs which contain sequences adjacent to the 5' ESTs. The extended cDNAs may include the entire coding sequence of the protein encoded by the corresponding mRNA, including the authentic translation start site, the signal sequence, and the sequence encoding the mature protein remaining after cleavage of the signal peptide. Such extended cDNAs are referred to herein as "full length cDNAs." Alternatively, the extended cDNAs may include only the sequence encoding the mature protein remaining after cleavage of the signal peptide, or only the sequence encoding the signal peptide.

5

10

15

20

25

Example 27 below describes a general method for obtaining extended cDNAs using 5' ESTs. Example 28 below provides experimental results, using the method explained in example 27, describing several extended cDNAs including the entire coding sequence and authentic 5' end of the corresponding mRNA for several secreted proteins.

The methods of Examples 27, 28, and 29 can also be used to obtain extended cDNAs which encode less than the entire coding sequence of the secreted proteins encoded by the genes corresponding to the 5' ESTs. In some embodiments, the extended cDNAs isolated using these methods encode at least 10 amino acids of one of the proteins encoded by the sequences of SEQ ID NOs: 38-184. In further embodiments, the extended cDNAs encode at least 20 amino acids of the proteins encoded by the sequences of SEQ ID NOs: 38-184. In further embodiments, the extended cDNAs encode at least 30 amino amino acids of the sequences of SEQ ID NOs: 38-184. In a preferred embodiment, the extended cDNAs encode a full length protein sequence, which includes the protein coding sequences of SEQ ID NOs: 38-184.

EXAMPLE 27

General Method for Using 5' ESTs to Clone and Sequence cDNAs which Include the Entire Coding Region and the Authentic 5' End of the Corresponding mRNA

The following general method has been used to quickly and efficiently isolate extended cDNAs having the authentic 5' ends of their corresponding mRNAs as well as

the full protein coding sequence and including sequence adjacent to the sequences of the 5' ESTs used to obtain them. This method may be applied to obtain extended cDNAs for any 5' EST in the NetGeneTM database, including those 5' ESTs encoding polypeptides belonging to secreted proteins. The method is summarized in figure 3.

5

10

15

20

25

1. Obtention of Extended cDNAs

a) First strand synthesis

The method takes advantage of the known 5' sequence of the mRNA. A reverse transcription reaction is conducted on purified mRNA with a poly 14dT primer containing a 49 nucleotide sequence at its 5' end allowing the addition of a known sequence at the end of the cDNA which corresponds to the 3' end of the mRNA. For example, the primer may have the following sequence: 5'-ATC GTT GAG ACT CGT ACC AGC AGA GTC ACG AGA GAG ACT ACA CGG TAC TGG TTT TTT TTT TTT TTVN -3' (SEQ ID NO:14). Those skilled in the art will appreciate that other sequences may also be added to the poly dT sequence and used to prime the first strand synthesis. Using this primer and a reverse transcriptase such as the Superscript II (Gibco BRL) or Rnase H Minus M-MLV (Promega) enzyme, a reverse transcript anchored at the 3' polyA site of the RNAs is generated.

After removal of the mRNA hybridized to the first cDNA strand by alkaline hydrolysis, the products of the alkaline hydrolysis and the residual poly dT primer are eliminated with an exclusion column such as an AcA34 (Biosepra) matrix as explained in Example 11.

b) Second strand synthesis

A pair of nested primers on each end is designed based on the known 5' sequence from the 5' EST and the known 3' end added by the poly dT primer used in the first strand synthesis. Softwares used to design primers are either based on GC content and melting temperatures of oligonucleotides, such as OSP (Illier and Green, PCR Meth. Appl. 1:124-128, 1991), or based on the octamer frequency disparity method (Griffais et al., Nucleic Acids Res. 19: 3887-3891, 1991) such as PC-Rare (http://bioinformatics.weizmann.ac.il/software/PC-Rare/doc/manuel.html).

10

15

20

25

Preferably, the nested primers at the 5' end are separated from one another by four to nine bases. The 5' primer sequences may be selected to have melting temperatures and specificities suitable for use in PCR

Preferably, the nested primers at the 3' end are separated from one another by four to nine bases. For example, the nested 3' primers may have the following sequences: (5'- CCA GCA GAG TCA CGA GAG AGA CTA CAC GG -3'(SEQ ID NO:15), and 5'- CAC GAG AGA GAC TAC ACG GTA CTG G -3' (SEQ ID NO:16). These primers were selected because they have melting temperatures and specificities compatible with their use in PCR. However, those skilled in the art will appreciate that other sequences may also be used as primers.

The first PCR run of 25 cycles is performed using the Advantage Tth Polymerase Mix (Clontech) and the outer primer from each of the nested pairs. A second 20 cycle PCR using the same enzyme and the inner primer from each of the nested pairs is then performed on 1/2500 of the first PCR product. Thereafter, the primers and nucleotides are removed.

2. Sequencing of Full Length Extended cDNAs or Fragments Thereof

Due to the lack of position constraints on the design of 5' nested primers compatible for PCR use using the OSP software, amplicons of two types are obtained. Preferably, the second 5' primer is located upstream of the translation initiation codon thus yielding a nested PCR product containing the whole coding sequence. Such a full length extended cDNA undergoes a direct cloning procedure as described in section a. However, in some cases, the second 5' primer is located downstream of the translation initiation codon, thereby yielding a PCR product containing only part of the ORF. Such incomplete PCR products are submitted to a modified procedure described in section b. a) Nested PCR products containing complete ORFs

When the resulting nested PCR product contains the complete coding sequence, as

predicted from the 5'EST sequence, it is cloned in an appropriate vector such as pED6dpc2, as described in section 3.

30 b) Nested PCR products containing incomplete ORFs

10

15

20

25

30

When the amplicon does not contain the complete coding sequence, intermediate steps are necessary to obtain both the complete coding sequence and a PCR product containing the full coding sequence. The complete coding sequence can be assembled from several partial sequences determined directly from different PCR products as described in the following section.

Once the full coding sequence has been completely determined, new primers compatible for PCR use are designed to obtain amplicons containing the whole coding region. However, in such cases, 3' primers compatible for PCR use are located inside the 3' UTR of the corresponding mRNA, thus yielding amplicons which lack part of this region, *i.e.* the polyA tract and sometimes the polyadenylation signal, as illustrated in figure 3. Such full length extended cDNAs are then cloned into an appropriate vector as described in section 3.

c) Sequencing extended cDNAs

Sequencing of extended cDNAs is performed using a Die Terminator approach with the AmpliTaq DNA polymerase FS kit available from Perkin Elmer.

In order to sequence PCR fragments, primer walking is performed using software such as OSP to choose primers and automated computer software such as ASMG (Sutton et al., Genome Science Technol. 1: 9-19, 1995) to construct contigs of walking sequences including the initial 5' tag using minimum overlaps of 32 nucleotides. Preferably, primer walking is performed until the sequences of full length cDNAs are obtained.

Completion of the sequencing of a given extended cDNA fragment is assessed as follows. Since sequences located after a polyA tract are difficult to determine precisely in the case of uncloned products, sequencing and primer walking processes for PCR products are interrupted when a polyA tract is identified in extended cDNAs obtained as described in case b. The sequence length is compared to the size of the nested PCR product obtained as described above. Due to the limited accuracy of the determination of the PCR product size by gel electrophoresis, a sequence is considered complete if the size of the obtained sequence is at least 70 % the size of the first nested PCR product. If the length of the sequence determined from the computer analysis is not at least 70% of the length of the nested PCR product, these PCR products are cloned and the sequence of the insertion is determined. When Northern blot data are available, the size of the mRNA detected for a given PCR

47

product is used to finally assess that the sequence is complete. Sequences which do not fulfill the above criteria are discarded and will undergo a new isolation procedure.

Sequence data of all extended cDNAs are then transferred to a proprietary database, where quality controls and validation steps are carried out as described in example 15.

3. Cloning of Full Length Extended cDNAs

5

10

15

20

25

30

The PCR product containing the full coding sequence is then cloned in an appropriate vector. For example, the extended cDNAs can be cloned into the expression vector pED6dpc2 (DiscoverEase, Genetics Institute, Cambridge, MA) as follows. pED6dpc2 vector DNA is prepared with blunt ends by performing an EcoRI digestion followed by a fill in reaction. The blunt ended vector is dephosphorylated. After removal of PCR primers and ethanol precipitation, the PCR product containing the full coding sequence or the extended cDNA obtained as described above is phosphorylated with a kinase subsequently removed by phenol-Sevag extraction and precipitation. The double stranded extended cDNA is then ligated to the vector and the resulting expression plasmid introduced into appropriate host cells.

Since the PCR products obtained as described above are blunt ended molecules that can be cloned in either direction, the orientation of several clones for each PCR product is determined. Then, 4 to 10 clones are ordered in microtiter plates and subjected to a PCR reaction using a first primer located in the vector close to the cloning site and a second primer located in the portion of the extended cDNA corresponding to the 3' end of the mRNA. This second primer may be the antisense primer used in anchored PCR in the case of direct cloning (case a) or the antisense primer located inside the 3'UTR in the case of indirect cloning (case b). Clones in which the start codon of the extended cDNA is operably linked to the promoter in the vector so as to permit expression of the protein encoded by the extended cDNA are conserved and sequenced. In addition to the ends of cDNA inserts, approximately 50 bp of vector DNA on each side of the cDNA insert are also sequenced.

The cloned PCR products are then entirely sequenced according to the aforementioned procedure. In this case, contigation of long fragments is then performed on walking sequences that have already contigated for uncloned PCR products during

primer walking. Sequencing of cloned amplicons is complete when the resulting contigs include the whole coding region as well as overlapping sequences with vector DNA on both ends.

5 4. Computer analysis of Full Length Extended cDNA

Sequences of all full length extended cDNAs are then submitted to further analysis as described below. Before searching the extended full length cDNAs for sequences of interest, extended cDNAs which are not of interest (vector RNAs, transfer RNAs, ribosomal RNAs, mitochondrial RNAs, prokaryotic RNAs and fungal RNAs) are discarded using methods essentially similar to those described for 5'ESTs in Example 18.

a) Identification of structural features

10

15

20

25

30

Structural features, e.g. polyA tail and polyadenylation signal, of the sequences of full length extended cDNAs are subsequently determined as follows.

A polyA tail is defined as a homopolymeric stretch of at least 11 A with at most one alternative base within it. The polyA tail search is restricted to the last 100 nt of the sequence and limited to stretches of 11 consecutive A's because sequencing reactions are often not readable after such a polyA stretch. Stretches having more than 90% homology over 8 nucleotides are identified as polyA tails using BLAST2N.

To search for a polyadenylation signal, the polyA tail is clipped from the full-length sequence. The 50 bp preceding the polyA tail are first searched for the canonic polyadenylation AAUAAA signal and, if the canonic signal is not detected, for the alternative AUUAAA signal (Sheets et al., Nuc. Acids Res. 18: 5799-5805, 1990). If neither of these consensus polyadenylation signals is found, the canonic motif is searched again allowing one mismatch to account for possible sequencing errors. More than 85 % of identified polyadenylation signals of either type actually ends 10 to 30 bp from the polyA tail. Alternative AUUAAA signals represents approximately 15 % of the total number of identified polyadenylation signals.

b) Identification of functional features

Functional features, e.g. ORFs and signal sequences, of the sequences of full length extended cDNAs were subsequently determined as follows.

The 3 upper strand frames of extended cDNAs are searched for ORFs defined as the maximum length fragments beginning with a translation intiation codon and ending with a stop codon. ORFs encoding at least 20 amino acids are preferred.

Each found ORF is then scanned for the presence of a signal peptide in the first 50 amino-acids or, where appropriate, within shorter regions down to 20 amino acids or less in the ORF, using the matrix method of von Heijne (*Nuc. Acids Res.* 14: 4683-4690, 1986), the disclosure of which is incorporated herein by reference as described in Example 22.

c) Homology to either nucleotidic or proteic sequences

5

10

15

20

25

30

Categorization of full-length sequences may be achieved using procedures essentially similar to those described for 5'ESTs in Example 24.

Extended cDNAs prepared as described above may be subsequently engineered to obtain nucleic acids which include desired portions of the extended cDNA using conventional techniques such as subcloning, PCR, or *in vitro* oligonucleotide synthesis. For example, nucleic acids which include only the full coding sequences (*i.e.* the sequences encoding the signal peptide and the mature protein remaining after the signal peptide is cleaved off) may be obtained using techniques known to those skilled in the art. Alternatively, conventional techniques may be applied to obtain nucleic acids which contain only the coding sequences for the mature protein remaining after the signal peptide is cleaved off or nucleic acids which contain only the coding sequences for the signal peptides.

Similarly, nucleic acids containing any other desired portion of the coding sequences for the secreted protein may be obtained. For example, the nucleic acid may contain at least 10 consecutive bases of an extended cDNA such as one of the extended cDNAs described below. In another embodiment, the nucleic acid may contain at least 15 consecutive bases of an extended cDNA such as one of the extended cDNAs described below. Alternatively, the nucleic acid may contain at least 20 consecutive bases of an extended cDNA such as one of the extended cDNAs described below. In another embodiment, the nucleic acid may contain at least 25 consecutive bases of an extended cDNAs uch as one of the extended cDNAs described below. In yet another embodiment, the nucleic acid may contain at least 40 described below. In yet another embodiment, the nucleic acid may contain at least 40

50

consecutive bases of an extended cDNA such as one of the extended cDNAs described below.

Once an extended cDNA has been obtained, it can be sequenced to determine the amino acid sequence it encodes. Once the encoded amino acid sequence has been determined, one can create and identify any of the many conceivable cDNAs that will encode that protein by simply using the degeneracy of the genetic code. For example, allelic variants or other homologous nucleic acids can be identified as described below. Alternatively, nucleic acids encoding the desired amino acid sequence can be synthesized *in vitro*.

In a preferred embodiment, the coding sequence may be selected using the known codon or codon pair preferences for the host organism in which the cDNA is to be expressed.

The extended cDNAs derived from the 5' ESTS of the present invention were obtained as described in Example 28 below.

EXAMPLE 28

15

20

25

30

10

5

Characterization of cloned extended cDNAs obtained using 5' ESTs

The procedure described in Example 27 above was used to obtain the extended cDNAs derived from the 5' ESTs of the present invention in a variety of tissues. The following list provides a few examples of thus obtained extended cDNAs.

Using this approach, the full length cDNA of SEQ ID NO:17 (internal identification number 48-19-3-G1-FL1) was obtained. This cDNA falls into the "EST-ext" category described above and encodes the signal peptide MKKVLLLITAILAVAVG (SEQ ID NO: 18) having a von Heijne score of 8.2.

The full length cDNA of SEQ ID NO:19 (internal identification number 58-34-2-E7-FL2) was also obtained using this procedure. This cDNA falls into the "EST-ext" category described above and encodes the signal peptide MWWFQQGLSFLPSALVIWTSA (SEQ ID NO:20) having a von Heijne score of 5.5.

Another full length cDNA obtained using the procedure described above has the sequence of SEQ ID NO:21 (internal identification number 51-27-1-E8-FL1). This cDNA, falls into the "EST-ext" category described above and encodes the signal peptide MVLTTLPSANSANSPVNMPTTGPNSLSYASSALSPCLT (SEQ ID NO:22) having a von Heijne score of 5.9.

The above procedure was also used to obtain a full length cDNA having the sequence of SEQ ID NO:23 (internal identification number 76-4-1-G5-FL1). This cDNA falls into the "EST-ext" category described above and encodes the signal peptide ILSTVTALTFAXA (SEQ ID NO:24) having a von Heijne score of 5.5.

The full length cDNA of SEQ ID NO:25 (internal identification number 51-3-3-B10-FL3) was also obtained using this procedure. This cDNA falls into the "new" category described above and encodes a signal peptide LVLTLCTLPLAVA (SEQ ID NO:26) having a von Heijne score of 10.1.

5

10

15

20

25

30

The full length cDNA of SEQ ID NO:27 (internal identification number 58-35-2-F10-FL2) was also obtained using this procedure. This cDNA falls into the "new" category described above and encodes a signal peptide LWLLFFLVTAIHA (SEQ ID NO:28) having a von Heijne score of 10.7.

Bacterial clones containing plasmids containing the full length cDNAs described above are presently stored in the inventor's laboratories under the internal identification numbers provided above. The inserts may be recovered from the stored materials by growing an aliquot of the appropriate bacterial clone in the appropriate medium. The plasmid DNA can then be isolated using plasmid isolation procedures familiar to those skilled in the art such as alkaline lysis minipreps or large scale alkaline lysis plasmid isolation procedures. If desired the plasmid DNA may be further enriched by centrifugation on a cesium chloride gradient, size exclusion chromatography, or anion exchange chromatography. The plasmid DNA obtained using these procedures may then be manipulated using standard cloning techniques familiar to those skilled in the art. Alternatively, a PCR can be done with primers designed at both ends of the cDNA insertion. The PCR product which corresponds to the cDNA can then be manipulated using standard cloning techniques familiar to those skilled in the art.

The polypeptides encoded by the extended cDNAs may be screened for the presence of known structural or functional motifs or for the presence of signatures, small amino acid sequences which are well conserved amongst the members of a protein family. The conserved regions have been used to derive consensus patterns or matrices included in the PROSITE data bank, in particular in the file prosite.dat (Release 13.0 of November 1995, located at http://expasy.hcuge.ch/sprot/prosite.html. Prosite_convert and prosite_scan

programs (http://ulrec3.unil.ch/ftpserveur/prosite_scan) may be used to find signatures on the extended cDNAs.

For each pattern obtained with the prosite_convert program from the prosite.dat file, the accuracy of the detection on a new protein sequence may be assessed by evaluating the frequency of irrelevant hits on the population of human secreted proteins included in the data bank SWISSPROT. The ratio between the number of hits on shuffled proteins (with a window size of 20 amino acids) and the number of hits on native (unshuffled) proteins may be used as an index. Every pattern for which the ratio is greater than 20% (one hit on shuffled proteins for 5 hits on native proteins) may be skipped during the search with prosite_scan. The program used to shuffle protein sequences (db_shuffled) and the program used to determine the statistics for each pattern in the protein data banks (prosite_statistics) are available on the ftp site http://ulrec3.unil.ch/ftpserveur/prosite_scan.

In addition to PCR based methods for obtaining extended cDNAs, traditional hybridization based methods may also be employed. These methods may also be used to obtain the genomic DNAs which encode the mRNAs from which the 5' ESTs were derived, mRNAs corresponding to the extended cDNAs, or nucleic acids which are homologous to extended cDNAs or 5' ESTs. Example 29 below provides examples of such methods.

20 EXAMPLE 29

5

10

15

25

30

Methods for Obtaining cDNAs which include the Entire Coding Region and the Authentic 5'End of the Corresponding mRNA

A full length cDNA library can be made using the strategies described in Examples 13, 14, 15, and 16 above by replacing the random nonamer used in Example 14 with an oligo-dT primer. For instance, the oligonucleotide of SEQ ID NO:14 may be used.

Alternatively, a cDNA library or genomic DNA library may be obtained from a commercial source or made using techniques familiar to those skilled in the art. Such cDNA or genomic DNA librairies may be used to isolate extended cDNAs obtained from 5' EST or nucleic acids homologous to extended cDNAs or 5' EST as follows. The cDNA library or genomic DNA library is hybridized to a detectable probe comprising at least 10 consecutive nucleotides from the 5' EST or extended cDNA using conventional techniques. Preferably,

the probe comprises at least 12, 15, or 17 consecutive nucleotides from the 5' EST or extended cDNA. More preferably, the probe comprises at least 20 to 30 consecutive nucleotides from the 5' EST or extended cDNA. In some embodiments, the probe comprises more than 30 nucleotides from the 5' EST or extended cDNA.

Techniques for identifying cDNA clones in a cDNA library which hybridize to a given probe sequence are disclosed in Sambrook et al., Molecular Cloning: A Laboratory Manual 2d Ed., Cold Spring Harbor Laboratory Press, 1989, the disclosure of which is incorporated herein by reference. The same techniques may be used to isolate genomic DNAs.

5

10

15

20

25

Briefly, cDNA or genomic DNA clones which hybridize to the detectable probe are identified and isolated for further manipulation as follows. A probe comprising at least 10 consecutive nucleotides from the 5' EST or extended cDNA is labeled with a detectable label such as a radioisotope or a fluorescent molecule. Preferably, the probe comprises at least 12, 15, or 17 consecutive nucleotides from the 5' EST or extended cDNA. More preferably, the probe comprises 20 to 30 consecutive nucleotides from the 5' EST or extended cDNA. In some embodiments, the probe comprises more than 30 nucleotides from the 5' EST or extended cDNA.

Techniques for labeling the probe are well known and include phosphorylation with polynucleotide kinase, nick translation, *in vitro* transcription, and non radioactive techniques. The cDNAs or genomic DNAs in the library are transferred to a nitrocellulose or nylon filter and denatured. After blocking of non specific sites, the filter is incubated with the labeled probe for an amount of time sufficient to allow binding of the probe to cDNAs or genomic DNAs containing a sequence capable of hybridizing thereto.

By varying the stringency of the hybridization conditions used to identify extended cDNAs or genomic DNAs which hybridize to the detectable probe, extended cDNAS having different levels of homology to the probe can be identified and isolated as described below.

1. Identification of Extended cDNA or Genomic cDNA Sequences Having a High Degree of Homology to the Labeled Probe

10

15

20

25

30

To identify extended cDNAs or genomic DNAs having a high degree of homology to the probe sequence, the melting temperature of the probe may be calculated using the following formulas:

For probes between 14 and 70 nucleotides in length the melting temperature (Tm) is calculated using the formula: Tm=81.5+16.6(log [Na+])+0.41(fraction G+C)-(600/N) where N is the length of the probe.

If the hybridization is carried out in a solution containing formamide, the melting temperature may be calculated using the equation Tm=81.5+16.6(log [Na+])+0.41(fraction G+C)-(0.63% formamide)-(600/N) where N is the length of the probe.

Prehybridization may be carried out in 6X SSC, 5X Denhardt's reagent, 0.5% SDS, 100 µg denatured fragmented salmon sperm DNA or 6X SSC, 5X Denhardt's reagent, 0.5% SDS, 100 µg denatured fragmented salmon sperm DNA, 50% formamide. The formulas for SSC and Denhardt's solutions are listed in Sambrook *et al.*, *supra*.

Hybridization is conducted by adding the detectable probe to the prehybridization solutions listed above. Where the probe comprises double stranded DNA, it is denatured before addition to the hybridization solution. The filter is contacted with the hybridization solution for a sufficient period of time to allow the probe to hybridize to extended cDNAs or genomic DNAs containing sequences complementary thereto or homologous thereto. For probes over 200 nucleotides in length, the hybridization may be carried out at 15-25°C below the Tm. For shorter probes, such as oligonucleotide probes, the hybridization may be conducted at 15-25°C below the Tm. Preferably, for hybridizations in 6X SSC, the hybridization is conducted at approximately 68°C. Preferably, for hybridizations in 50% formamide containing solutions, the hybridization is conducted at approximately 42°C.

All of the foregoing hybridizations would be considered to be under "stringent" conditions.

Following hybridization, the filter is washed in 2X SSC, 0.1% SDS at room temperature for 15 minutes. The filter is then washed with 0.1X SSC, 0.5% SDS at room temperature for 30 minutes to 1 hour. Thereafter, the solution is washed at the hybridization temperature in 0.1X SSC, 0.5% SDS. A final wash is conducted in 0.1X SSC at room temperature.

Extended cDNAs, nucleic acids homologous to extended cDNAs or 5' ESTs, or genomic DNAs which have hybridized to the probe are identified by autoradiography or other conventional techniques.

5 <u>2. Obtention of Extended cDNA or Genomic cDNA Sequences Having Lower Degrees</u> of Homology to the Labeled Probe

The above procedure may be modified to identify extended cDNAs, nucleic acids homologous to extended cDNAs, or genomic DNAs having decreasing levels of homology to the probe sequence. For example, to obtain extended cDNAs, nucleic acids homologous to extended cDNAs, or genomic DNAs of decreasing homology to the detectable probe, less stringent conditions may be used. For example, the hybridization temperature may be decreased in increments of 5°C from 68°C to 42°C in a hybridization buffer having a sodium concentration of approximately 1M. Following hybridization, the filter may be washed with 2X SSC, 0.5% SDS at the temperature of hybridization. These conditions are considered to be "moderate" conditions above 50°C and "low" conditions below 50°C.

10

15

20

30

Alternatively, the hybridization may be carried out in buffers, such as 6X SSC, containing formamide at a temperature of 42°C. In this case, the concentration of formamide in the hybridization buffer may be reduced in 5% increments from 50% to 0% to identify clones having decreasing levels of homology to the probe. Following hybridization, the filter may be washed with 6X SSC, 0.5% SDS at 50°C. These conditions are considered to be "moderate" conditions above 25% formamide and "low" conditions below 25% formamide.

Extended cDNAs, nucleic acids homologous to extended cDNAs, or genomic DNAs which have hybridized to the probe are identified by autoradiography.

25 <u>3. Determination of the Degree of Homology Between the Obtained Extended cDNAs</u> and the Labeled Probe

If it is desired to obtain nucleic acids homologous to extended cDNAs, such as allelic variants thereof or nucleic acids encoding proteins related to the proteins encoded by the extended cDNAs, the level of homology between the hybridized nucleic acid and the extended cDNA or 5' EST used as the probe may be further determined using BLAST2N; parameters may be adapted depending on the sequence length and degree of homology

studied. To determine the level of homology between the hybridized nucleic acid and the extended cDNA or 5'EST from which the probe was derived, the nucleotide sequences of the hybridized nucleic acid and the extended cDNA or 5'EST from which the probe was derived are compared. For example, using the above methods, nucleic acids having at least 95% nucleic acid homology to the extended cDNA or 5'EST from which the probe was derived may be obtained and identified. Similarly, by using progressively less stringent hybridization conditions one can obtain and identify nucleic acids having at least 90%, at least 85%, at least 80% or at least 75% homology to the extended cDNA or 5'EST from which the probe was derived.

10

15

20

5

To determine whether a clone encodes a protein having a given amount of homology to the protein encoded by the extended cDNA or 5' EST, the amino acid sequence encoded by the extended cDNA or 5' EST is compared to the amino acid sequence encoded by the hybridizing nucleic acid. Homology is determined to exist when an amino acid sequence in the extended cDNA or 5' EST is closely related to an amino acid sequence in the hybridizing nucleic acid. A sequence is closely related when it is identical to that of the extended cDNA or 5' EST or when it contains one or more amino acid substitutions therein in which amino acids having similar characteristics have been substituted for one another. Using the above methods and algorithms such as FASTA with parameters depending on the sequence length and degree of homology studied, one can obtain nucleic acids encoding proteins having at least 95%, at least 90%, at least 85%, at least 80% or at least 75% homology to the proteins encoded by the extended cDNA or 5'EST from which the probe was derived.

In addition to the above described methods, other protocols are available to obtain extended cDNAs using 5' ESTs as outlined in the following paragraphs.

25

Extended cDNAs may be prepared by obtaining mRNA from the tissue, cell, or organism of interest using mRNA preparation procedures utilizing polyA selection procedures or other techniques known to those skilled in the art. A first primer capable of hybridizing to the polyA tail of the mRNA is hybridized to the mRNA and a reverse transcription reaction is performed to generate a first cDNA strand.

30

The first cDNA strand is hybridized to a second primer containing at least 10 consecutive nucleotides of the sequences of SEQ ID NOs 38-184. Preferably, the primer

10

15

20

25

30

comprises at least 12, 15, or 17 consecutive nucleotides from the sequences of SEQ ID NOs 38-184. More preferably, the primer comprises 20 to 30 consecutive nucleotides from the sequences of SEQ ID NOs 38-184. In some embodiments, the primer comprises more than 30 nucleotides from the sequences of SEQ ID NOs 38-184. If it is desired to obtain extended cDNAs containing the full protein coding sequence, including the authentic translation initiation site, the second primer used contains sequences located upstream of the translation initiation site. The second primer is extended to generate a second cDNA strand complementary to the first cDNA strand. Alternatively, RT-PCR may be performed as described above using primers from both ends of the cDNA to be obtained.

Extended cDNAs containing 5' fragments of the mRNA may be prepared by hybridizing an mRNA comprising the sequence of the 5'EST for which an extended cDNA is desired with a primer comprising at least 10 consecutive nucleotides of the sequences complementary to the 5'EST and reverse transcribing the hybridized primer to make a first cDNA strand from the mRNAs. Preferably, the primer comprises at least 12, 15, or 17 consecutive nucleotides from the 5'EST. More preferably, the primer comprises 20 to 30 consecutive nucleotides from the 5'EST.

Thereafter, a second cDNA strand complementary to the first cDNA strand is synthesized. The second cDNA strand may be made by hybridizing a primer complementary to sequences in the first cDNA strand to the first cDNA strand and extending the primer to generate the second cDNA strand.

The double stranded extended cDNAs made using the methods described above are isolated and cloned. The extended cDNAs may be cloned into vectors such as plasmids or viral vectors capable of replicating in an appropriate host cell. For example, the host cell may be a bacterial, mammalian, avian, or insect cell.

Techniques for isolating mRNA, reverse transcribing a primer hybridized to mRNA to generate a first cDNA strand, extending a primer to make a second cDNA strand complementary to the first cDNA strand, isolating the double stranded cDNA and cloning the double stranded cDNA are well known to those skilled in the art and are described in Current Protocols in Molecular Biology, John Wiley and Sons, Inc. 1997 and Sambrook et al., Molecular Cloning: A Laboratory Manual, Second Edition, Cold Spring Harbor Laboratory Press, 1989, the entire disclosures of which are incorporated herein by reference.

Alternatively, procedures such as the one described in Example 29 may be used for obtaining full length cDNAs or extended cDNAs. In this approach, full length or extended cDNAs are prepared from mRNA and cloned into double stranded phagemids as follows. The cDNA library in the double stranded phagemids is then rendered single stranded by treatment with an endonuclease, such as the Gene II product of the phage F1, and an exonuclease (Chang et al., Gene 127:95-8, 1993). A biotinylated oligonucleotide comprising the sequence of a 5' EST, or a fragment containing at least 10 nucleotides thereof, is hybridized to the single stranded phagemids. Preferably, the fragment comprises at least 12, 15, or 17 consecutive nucleotides from the 5' EST. More preferably, the fragment comprises 20-30 consecutive nucleotides from the 5' EST. In some procedures, the fragment may comprise more than 30 consecutive nucleotides from the 5' EST.

Hybrids between the biotinylated oligonucleotide and phagemids having inserts containing the 5' EST sequence are isolated by incubating the hybrids with streptavidin coated paramagnetic beads and retrieving the beads with a magnet (Fry et al., Biotechniques, 13: 124-131, 1992). Therafter, the resulting phagemids containing the 5' EST sequence are released from the beads and converted into double stranded DNA using a primer specific for the 5' EST sequence. Alternatively, protocoles such as the Gene Trapper kit (Gibco BRL) may be used. The resulting double stranded DNA is transformed into bacteria. Extended cDNAs containing the 5' EST sequence are identified by colony PCR or colony hybridization.

20

25

30

5

10

15

Using any of the above described methods in section III, a plurality of extended cDNAs containing full length protein coding sequences or sequences encoding only the mature protein remaining after the signal peptide is cleaved off may be provided as cDNA libraries for subsequent evaluation of the encoded proteins or use in diagnostic assays as described below.

IV. Expression of Proteins Encoded by Extended cDNAs Isolated Using 5' ESTs

Extended cDNAs containing the full protein coding sequences of their corresponding mRNAs or portions thereof, such as cDNAs encoding the mature protein, may be used to express the encoded secreted proteins or portions thereof as described in Example 30 below. If desired, the extended cDNAs may contain the sequences encoding the signal peptide to

59

facilitate secretion of the expressed protein. It will be appreciated that a plurality of extended cDNAs containing the full protein coding sequences or portions thereof may be simultaneously cloned into expression vectors to create an expression library for analysis of the encoded proteins as described below.

5

10

15

20

25

30

EXAMPLE 30

Expression of the Proteins Encoded by the Genes Corresponding to 5'ESTS or Portions Thereof

To express the proteins encoded by the genes corresponding to 5' ESTs (or portions thereof), full length cDNAs containing the entire protein coding region or extended cDNAs containing sequences adjacent to the 5' ESTs (or portions thereof) are obtained as described in Examples 27-29 and cloned into a suitable expression vector. If desired, the nucleic acids may contain the sequences encoding the signal peptide to facilitate secretion of the expressed protein. The nucleic acids inserted into the expression vectors may also contain sequences upstream of the sequences encoding the signal peptide, such as sequences which regulate expression levels or sequences which confer tissue specific expression.

The nucleic acid encoding the protein or polypeptide to be expressed is operably linked to a promoter in an expression vector using conventional cloning technology. The expression vector may be any of the mammalian, yeast, insect or bacterial expression systems known in the art. Commercially available vectors and expression systems are available from a variety of suppliers including Genetics Institute (Cambridge, MA), Stratagene (La Jolla, California), Promega (Madison, Wisconsin), and Invitrogen (San Diego, California). If desired, to enhance expression and facilitate proper protein folding, the codon context and codon pairing of the sequence may be optimized for the particular expression organism in which the expression vector is introduced, as explained by Hatfield, *et al.*, U.S. Patent No. 5,082,767, incorporated herein by this reference.

The cDNA cloned into the expression vector may encode the entire protein (*i.e.* the signal peptide and the mature protein), the mature protein (*i.e.* the protein created by cleaving the signal peptide off), only the signal peptide or any other portion thereof.

The following is provided as one exemplary method to express the proteins encoded by the extended cDNAs corresponding to the 5' ESTs or the nucleic acids described above.

First, the methionine initiation codon for the gene and the polyA signal of the gene are identified. If the nucleic acid encoding the polypeptide to be expressed lacks a methionine to serve as the initiation site, an initiating methionine can be introduced next to the first codon of the nucleic acid using conventional techniques. Similarly, if the extended cDNA lacks a polyA signal, this sequence can be added to the construct by, for example, splicing out the polyA signal from pSG5 (Stratagene) using BgIII and SalI restriction endonuclease enzymes and incorporating it into the mammalian expression vector pXT1 (Stratagene). pXT1 contains the LTRs and a portion of the gag gene from Moloney Murine Leukemia Virus. The position of the LTRs in the construct allow efficient stable transfection. The vector includes the Herpes Simplex thymidine kinase promoter and the selectable neomycin gene. The extended cDNA or portion thereof encoding the polypeptide to be expressed is obtained by PCR from the bacterial vector using oligonucleotide primers complementary to the extended cDNA or portion thereof and containing restriction endonuclease sequences for Pst I incorporated into the 5'primer and BglII at the 5' end of the corresponding cDNA 3' primer, taking care to ensure that the extended cDNA is positioned with the poly A signal. The purified fragment obtained from the resulting PCR reaction is digested with PstI, blunt ended with an exonuclease, digested with Bgl II, purified and ligated to pXT1 containing a poly A signal and prepared for this ligation (blunt/BgIII).

5

10

15

20

25

30

The ligated product is transfected into mouse NIH 3T3 cells using Lipofectin (Life Technologies, Inc., Grand Island, New York) under conditions outlined in the product specification. Positive transfectants are selected after growing the transfected cells in 600 µg/ml G418 (Sigma, St. Louis, Missouri). Preferably the expressed protein is released into the culture medium, thereby facilitating purification.

Alternatively, the extended cDNAs may be cloned into pED6dpc2 as described above. The resulting pED6dpc2 constructs may be transfected into a suitable host cell, such as COS 1 cells. Methotrexate resistant cells are selected and expanded. Preferably, the protein expressed from the extended cDNA is released into the culture medium thereby facilitating purification.

Proteins in the culture medium are separated by gel electrophoresis. If desired, the proteins may be ammonium sulfate precipitated or separated based on size or charge prior to electrophoresis.

61

As a control, the expression vector lacking a cDNA insert is introduced into host cells or organisms and the proteins in the medium are harvested. The secreted proteins present in the medium are detected using techniques familiar to those skilled in the art such as Coomassie blue or silver staining or using antibodies against the protein encoded by the extended cDNA.

Antibodies capable of specifically recognizing the protein of interest may be generated using synthetic 15-mer peptides having a sequence encoded by the appropriate 5' EST, extended cDNA, or portion thereof. The synthetic peptides are injected into mice to generate antibody to the polypeptide encoded by the 5' EST, extended cDNA, or portion thereof.

Secreted proteins from the host cells or organisms containing an expression vector which contains the extended cDNA derived from a 5' EST or a portion thereof are compared to those from the control cells or organism. The presence of a band in the medium from the cells containing the expression vector which is absent in the medium from the control cells indicates that the extended cDNA encodes a secreted protein. Generally, the band corresponding to the protein encoded by the extended cDNA will have a mobility near that expected based on the number of amino acids in the open reading frame of the extended cDNA. However, the band may have a mobility different than that expected as a result of modifications such as glycosylation, ubiquitination, or enzymatic cleavage.

Alternatively, if the protein expressed from the above expression vectors does not contain sequences directing its secretion, the proteins expressed from host cells containing an expression vector with an insert encoding a secreted protein or portion thereof can be compared to the proteins expressed in control host cells containing the expression vector without an insert. The presence of a band in samples from cells containing the expression vector with an insert which is absent in samples from cells containing the expression vector without an insert indicates that the desired protein or portion thereof is being expressed. Generally, the band will have the mobility expected for the secreted protein or portion thereof. However, the band may have a mobility different than that expected as a result of modifications such as glycosylation, ubiquitination, or enzymatic cleavage.

5

10

15

20

25

62

The protein encoded by the extended cDNA may be purified using standard immunochromatography techniques. In such procedures, a solution containing the secreted protein, such as the culture medium or a cell extract, is applied to a column having antibodies against the secreted protein attached to the chromatography matrix. The secreted protein is allowed to bind the immunochromatography column. Thereafter, the column is washed to remove non-specifically bound proteins. The specifically bound secreted protein is then released from the column and recovered using standard techniques.

5

10

15

20

25

30

If antibody production is not possible, the extended cDNA sequence or portion thereof may be incorporated into expression vectors designed for use in purification schemes employing chimeric polypeptides. In such strategies, the coding sequence of the extended cDNA or portion thereof is inserted in frame with the gene encoding the other half of the chimera. The other half of the chimera may be β -globin or a nickel binding polypeptide. A chromatography matrix having antibody to β -globin or nickel attached thereto is then used to purify the chimeric protein. Protease cleavage sites may be engineered between the β -globin gene or the nickel binding polypeptide and the extended cDNA or portion thereof. Thus, the two polypeptides of the chimera may be separated from one another by protease digestion.

One useful expression vector for generating β-globin chimerics is pSG5 (Stratagene), which encodes rabbit β-globin. Intron II of the rabbit β-globin gene facilitates splicing of the expressed transcript, and the polyadenylation signal incorporated into the construct increases the level of expression. These techniques as described are well known to those skilled in the art of molecular biology. Standard methods are published in methods texts such as Davis *et al.*, (*Basic Methods in Molecular Biology*, Davis, Dibner, and Battey, ed., Elsevier Press, NY, 1986) and many of the methods are available from Stratagene, Life Technologies, Inc., or Promega. Polypeptide may additionally be produced from the construct using *in vitro* translation systems such as the *In vitro* ExpressTM Translation Kit (Stratagene).

Following expression and purification of the secreted proteins encoded by the 5' ESTs, extended cDNAs, or fragments thereof, the purified proteins may be tested for the ability to bind to the surface of various cell types as described in Example 31 below. It will be appreciated that a plurality of proteins expressed from these cDNAs may be included in a

63

panel of proteins to be simultaneously evaluated for the activities specifically described below, as well as other biological roles for which assays for determining activity are available.

EXAMPLE 31

Analysis of Secreted Proteins to Determine Whether they Bind to the Cell Surface

The proteins encoded by the 5' ESTs, extended cDNAs, or fragments thereof are cloned into expression vectors such as those described in Example 30. The proteins are purified by size, charge, immunochromatography or other techniques familiar to those skilled in the art. Following purification, the proteins are labeled using techniques known to those skilled in the art. The labeled proteins are incubated with cells or cell lines derived from a variety of organs or tissues to allow the proteins to bind to any receptor present on the cell surface. Following the incubation, the cells are washed to remove non-specifically bound protein. The labeled proteins are detected by autoradiography. Alternatively, unlabeled proteins may be incubated with the cells and detected with antibodies having a detectable label, such as a fluorescent molecule, attached thereto.

Specificity of cell surface binding may be analyzed by conducting a competition analysis in which various amounts of unlabeled protein are incubated along with the labeled protein. The amount of labeled protein bound to the cell surface decreases as the amount of competitive unlabeled protein increases. As a control, various amounts of an unlabeled protein unrelated to the labeled protein is included in some binding reactions. The amount of labeled protein bound to the cell surface does not decrease in binding reactions containing increasing amounts of unrelated unlabeled protein, indicating that the protein encoded by the cDNA binds specifically to the cell surface.

As discussed above, secreted proteins have been shown to have a number of important physiological effects and, consequently, represent a valuable therapeutic resource. The secreted proteins encoded by the extended cDNAs or portions thereof made according to Examples 27-29 may be evaluated to determine their physiological activities as described below.

5

10

15

20

64

EXAMPLE 32

Assaying the Proteins Expressed from Extended cDNAs or Portions Thereof for Cytokine, Cell Proliferation or Cell Differentiation Activity

5

10

15

20

25

30

As discussed above, secreted proteins may act as cytokines or may affect cellular proliferation or differentiation. Many protein factors discovered to date, including all known cytokines, have exhibited activity in one or more factor dependent cell proliferation assays, and hence the assays serve as a convenient confirmation of cytokine activity. The activity of a protein encoded by the extended cDNAs is evidenced by any one of a number of routine factor dependent cell proliferation assays for cell lines including, without limitation, 32D, DA2, DA1G, T10, B9, B9/11, BaF3, MC9/G, M⁺ (preB M⁺), 2E8, RB5, DA1, 123, T1165, HT2, CTLL2, TF-1, Mo7c and CMK. The proteins encoded by the above extended cDNAs or portions thereof may be evaluated for their ability to regulate T cell or thymocyte proliferation in assays such as those described above or in the following references, which are incorporated herein by reference: *Current Protocols in Immunology*, Ed. by Coligan *et al.*, Greene Publishing Associates and Wiley-Interscience; Takai *et al. J. Immunol.* 137:3494-3500, 1986., Bertagnolli *et al.*, J. *Immunol.* 145:1706-1712, 1990., Bertagnolli *et al.*, Cell. *Immunol.* 133:327-341, 1991; Bertagnolli, *et al.*, J. *Immunol.* 149:3778-3783, 1992; Bowman *et al.*, J. *Immunol.* 152:1756-1761, 1994.

In addition, numerous assays for cytokine production and/or the proliferation of spleen cells, lymph node cells and thymocytes are known. These include the techniques disclosed in *Current Protocols in Immunology, supra* 1:3.12.1-3.12.14; and Schreiber In *Current Protocols in Immunology, supra* 1:6.8.1-6.8.8.

The proteins encoded by the cDNAs may also be assayed for the ability to regulate the proliferation and differentiation of hematopoietic or lymphopoietic cells. Many assays for such activity are familiar to those skilled in the art, including the assays in the following references, which are incorporated herein by reference: Bottomly et al., In Current Protocols in Immunology., supra. 1: 6.3.1-6.3.12,; deVries et al., J. Exp. Med. 173:1205-1211, 1991; Moreau et al., Nature 36:690-692, 1988; Greenberger et al., Proc. Natl. Acad. Sci. U.S.A. 80:2931-2938, 1983; Nordan, R., In Current Protocols in Immunology., supra. 1: 6.6.1-6.6.5; Smith et al., Proc. Natl. Acad. Sci. U.S.A. 83:1857-1861, 1986; Bennett et al., in

10

15

Current Protocols in Immunology supra 1: 6.15.1; Ciarletta et al., In Current Protocols in Immunology. supra 1: 6.13.1.

The proteins encoded by the cDNAs may also be assayed for their ability to regulate T-cell responses to antigens. Many assays for such activity are familiar to those skilled in the art, including the assays described in the following references, which are incorporated herein by reference: Chapter 3 (In Vitro Assays for Mouse Lymphocyte Function), Chapter 6 (Cytokines and Their Cellular Receptors) and Chapter 7, (Immunologic Studies in Humans) in Current Protocols in Immunology supra; Weinberger et al., Proc. Natl. Acad. Sci. USA 77:6091-6095, 1980; Weinberger et al., Eur. J. Immun. 11:405-411, 1981; Takai et al., J. Immunol. 137:3494-3500, 1986; Takai et al., J. Immunol. 140:508-512, 1988.

Those proteins which exhibit cytokine, cell proliferation, or cell differentiation activity may then be formulated as pharmaceuticals and used to treat clinical conditions in which induction of cell proliferation or differentiation is beneficial. Alternatively, as described in more detail below, genes encoding these proteins or nucleic acids regulating the expression of these proteins may be introduced into appropriate host cells to increase or decrease the expression of the proteins as desired.

EXAMPLE 33

20

25

30

Assaying the Proteins Expressed from Extended cDNAs or Portions Thereof for Activity as Immune System Regulators

The proteins encoded by the cDNAs may also be evaluated for their effects as immune regulators. For example, the proteins may be evaluated for their activity to influence thymocyte or splenocyte cytotoxicity. Numerous assays for such activity are familiar to those skilled in the art including the assays described in the following references, which are incorporated herein by reference: Chapter 3 (In Vitro Assays for Mouse Lymphocyte Function 3.1-3.19) and Chapter 7 (Immunologic studies in Humans) in Current Protocols in Immunology, Coligan et al., Eds, Greene Publishing Associates and Wiley-Interscience; Herrmann et al., Proc. Natl. Acad. Sci. USA 78:2488-2492, 1981; Herrmann et al., J. Immunol. 128:1968-1974, 1982; Handa et al., J. Immunol. 135:1564-1572, 1985; Takai et al., J. Immunol. 137:3494-3500, 1986; Takai et al., J. Immunol. 140:508-512, 1988;

10

15

20

25

30

Bowman et al., J. Virology 61:1992-1998; Bertagnolli et al., Cell. Immunol. 133:327-341, 1991; Brown et al., J. Immunol. 153:3079-3092, 1994.

The proteins encoded by the cDNAs may also be evaluated for their effects on T-cell dependent immunoglobulin responses and isotype switching. Numerous assays for such activity are familiar to those skilled in the art, including the assays disclosed in the following references, which are incorporated herein by reference: Maliszewski, *J. Immunol.* 144:3028-3033, 1990; Mond *et al.* in *Current Protocols in Immunology*, 1:3.8.1-3.8.16, *supra*.

The proteins encoded by the cDNAs may also be evaluated for their effect on immune effector cells, including their effect on Th1 cells and cytotoxic lymphocytes. Numerous assays for such activity are familiar to those skilled in the art, including the assays disclosed in the following references, which are incorporated herein by reference: Chapter 3 (*In Vitro* Assays for Mouse Lymphocyte Function 3.1-3.19) and Chapter 7 (Immunologic Studies in Humans) in *Current Protocols in Immunology, supra*; Takai et al., J. Immunol. 137:3494-3500, 1986; Takai et al., J. Immunol. 140:508-512, 1988; Bertagnolli et al., J. Immunol. 149:3778-3783, 1992.

The proteins encoded by the cDNAs may also be evaluated for their effect on dendritic cell mediated activation of naive T-cells. Numerous assays for such activity are familiar to those skilled in the art, including the assays disclosed in the following references, which are incorporated herein by reference: Guery et al., J. Immunol. 134:536-544, 1995, Inaba et al., J. Exp. Med. 173:549-559, 1991; Macatonia et al., J. Immunol. 154:5071-5079, 1995; Porgador et al.J. Exp. Med 182:255-260, 1995; Nair et al., J. Virol. 67:4062-4069, 1993; Huang et al., Science 264:961-965, 1994; Macatonia et al.J. Exp. Med 169:1255-1264, 1989; Bhardwaj et al., Journal of Clinical Investigation 94:797-807, 1994; and Inaba et al., J. Exp. Med 172:631-640, 1990.

The proteins encoded by the cDNAs may also be evaluated for their influence on the lifetime of lymphocytes. Numerous assays for such activity are familiar to those skilled in the art, including the assays disclosed in the following references, which are incorporated herein by reference: Darzynkiewicz et al., Cytometry 13:795-808, 1992; Gorczyca et al., Leukemia 7:659-670, 1993; Gorczyca et al., Cancer Res. 53:1945-1951, 1993; Itoh et al., Cell 66:233-243, 1991; Zacharchuk, J. Immunol. 145:4037-4045, 1990; Zamai et al., Cytometry 14:891-897, 1993; Gorczyca et al., Int. J. Oncol. 1:639-648, 1992.

The proteins encoded by the cDNAs may also be evaluated for their influence on early steps of T-cell commitment and development. Numerous assays for such activity are familiar to those skilled in the art, including without limitation the assays disclosed in the following references, which are incorporated herein by references: Antica et al., Blood 84:111-117, 1994; Fine et al., Cell. Immunol. 155:111-122, 1994; Galy et al., Blood 85:2770-2778, 1995; Toki et al., Proc. Nat. Acad Sci. USA 88:7548-7551, 1991.

Those proteins which exhibit activity as immune system regulators activity may then be formulated as pharmaceuticals and used to treat clinical conditions in which regulation of immune activity is beneficial. For example, the protein may be useful in the treatment of various immune deficiencies and disorders (including severe combined immunodeficiency), e.g., in regulating (up or down) growth and proliferation of T and/or B lymphocytes, as well as effecting the cytolytic activity of NK cells and other cell populations. These immune deficiencies may be genetic or be caused by viral (e.g., HIV) as well as bacterial or fungal infections, or may result from autoimmune disorders. More specifically, infectious diseases caused by viral, bacterial, fungal or other infection may be treatable using a protein encoded by extended cDNAs derived from the 5' ESTs of the present invention, including infections by HIV, hepatitis viruses, herpesviruses, mycobacteria, Leishmania spp., plamodium and various fungal infections such as candidiasis. Of course, in this regard, a protein encoded by extended cDNAs derived from the 5' ESTs of the present invention may also be useful where a boost to the immune system generally may be desirable, *i.e.*, in the treatment of cancer.

Alternatively, proteins encoded by extended cDNAs derived from the 5' ESTs of the present invention may be used in treatment of autoimmune disorders including, for example, connective tissue disease, multiple sclerosis, systemic lupus erythematosus, rheumatoid arthritis, autoimmune pulmonary inflammation, Guillain-Barre syndrome, autoimmune thyroiditis, insulin dependent diabetes mellitis, myasthenia gravis, graft-versus-host disease and autoimmune inflammatory eye disease. Such a protein encoded by extended cDNAs derived from the 5' ESTs of the present invention may also to be useful in the treatment of allergic reactions and conditions, such as asthma (particularly allergic asthma) or other respiratory problems. Other conditions, in which immune suppression is desired (including, for example, organ transplantation), may also be treatable using a protein encoded by extended cDNAs derived from the 5' ESTs of the present invention.

68

Using the proteins of the invention it may also be possible to regulate immune responses either up or down.

Down regulation may involve inhibiting or blocking an immune response already in progress or may involve preventing the induction of an immune response. The functions of activated T-cells may be inhibited by suppressing T cell responses or by inducing specific tolerance in T cells, or both. Immunosuppression of T cell responses is generally an active non-antigen-specific process which requires continuous exposure of the T cells to the suppressive agent. Tolerance, which involves inducing non-responsiveness or anergy in T cells, is distinguishable from immunosuppression in that it is generally antigen-specific and persists after the end of exposure to the tolerizing agent. Operationally, tolerance can be demonstrated by the lack of a T cell response upon reexposure to specific antigen in the absence of the tolerizing agent.

5

10

15

20

25

30

Down regulating or preventing one or more antigen functions (including without limitation B lymphocyte antigen functions, such as, for example, B7 costimulation), e.g., preventing high level lymphokine synthesis by activated T cells, will be useful in situations of tissue, skin and organ transplantation and in graft-versus-host disease (GVHD). For example, blockage of T cell function should result in reduced tissue destruction in tissue transplantation. Typically, in tissue transplants, rejection of the transplant is initiated through its recognition as foreign by T cells, followed by an immune reaction that destroys the transplant. The administration of a molecule which inhibits or blocks interaction of a B7 lymphocyte antigen with its natural ligand(s) on immune cells (such as a soluble, monomeric form of a peptide having B7-2 activity alone or in conjunction with a monomeric form of a peptide having an activity of another B lymphocyte antigen (e.g., B7-1, B7-3) or blocking antibody), prior to transplantation, can lead to the binding of the molecule to the natural ligand(s) on the immune cells without transmitting the corresponding costimulatory signal. Blocking B lymphocyte antigen function in this matter prevents cytokine synthesis by immune cells, such as T cells, and thus acts as an immunosuppressant. Moreover, the lack of costimulation may also be sufficient to anergize the T cells, thereby inducing tolerance in a subject. Induction of long-term tolerance by B lymphocyte antigen-blocking reagents may avoid the necessity of repeated administration of these blocking reagents. To achieve

69

sufficient immunosuppression or tolerance in a subject, it may also be necessary to block the function of a combination of B lymphocyte antigens.

The efficacy of particular blocking reagents in preventing organ transplant rejection or GVHD can be assessed using animal models that are predictive of efficacy in humans. Examples of appropriate systems which can be used include allogeneic cardiac grafts in rats and xenogeneic pancreatic islet cell grafts in mice, both of which have been used to examine the immunosuppressive effects of CTLA4Ig fusion proteins *in vivo* as described in Lenschow *et al.*, *Science* 257:789-792, 1992 and Turka *et al.*, *Proc. Natl. Acad. Sci USA*, 89:11102-11105, 1992. In addition, murine models of GVHD (see Paul ed., *Fundamental Immunology*, Raven Press, New York, 1989, pp. 846-847) can be used to determine the effect of blocking B lymphocyte antigen function *in vivo* on the development of that disease.

5

10

15

20

25

30

Blocking antigen function may also be therapeutically useful for treating autoimmune diseases. Many autoimmune disorders are the result of inappropriate activation of T cells that are reactive against self tissue and which promote the production of cytokines and autoantibodies involved in the pathology of the diseases. Preventing the activation of autoreactive T cells may reduce or eliminate disease symptoms. Administration of reagents which block costimulation of T cells by disrupting receptor/ligand interactions of B lymphocyte antigens can be used to inhibit T cell activation and prevent production of autoantibodies or T cell-derived cytokines which potentially involved in the disease process. Additionally, blocking reagents may induce antigen-specific tolerance of autoreactive T cells which could lead to long-term relief from the disease. The efficacy of blocking reagents in preventing or alleviating autoimmune disorders can be determined using a number of well-characterized animal models of human autoimmune diseases. Examples include murine experimental autoimmune encephalitis, systemic lupus erythmatosis in MRL/pr/pr mice or NZB hybrid mice, murine autoimmuno collagen arthritis, diabetes mellitus in OD mice and BB rats, and murine experimental myasthenia gravis (see Paul ed., supra, pp. 840-856).

Upregulation of an antigen function (preferably a B lymphocyte antigen function), as a means of up regulating immune responses, may also be useful in therapy. Upregulation of immune responses may involve either enhancing an existing immune response or eliciting an initial immune response as shown by the following examples. For instance, enhancing an immune response through stimulating B lymphocyte antigen function may be useful in cases

70

of viral infection. In addition, systemic viral diseases such as influenza, the common cold, and encephalitis might be alleviated by the administration of stimulatory form of B lymphocyte antigens systemically.

Alternatively, antiviral immune responses may be enhanced in an infected patient by removing T cells from the patient, costimulating the T cells *in vitro* with viral antigen-pulsed APCs either expressing a peptide encoded by extended cDNAs derived from the 5' ESTs of the present invention or together with a stimulatory form of a soluble peptide encoded by extended cDNAs derived from the 5' ESTs of the present invention and reintroducing the *in vitro* primed T cells into the patient. The infected cells would now be capable of delivering a costimulatory signal to T cells *in vivo*, thereby activating the T cells.

5

10

15

20

25

30

In another application, upregulation or enhancement of antigen function (preferably B lymphocyte antigen function) may be useful in the induction of tumor immunity. Tumor cells (e.g., sarcoma, melanoma, lymphoma, leukemia, neuroblastoma, carcinoma) transfected with a nucleic acid encoding at least one peptide encoded by extended cDNAs derived from the 5' ESTs of the present invention can be administered to a subject to overcome tumor-specific tolerance in the subject. If desired, the tumor cell can be transfected to express a combination of peptides. For example, tumor cells obtained from a patient can be transfected *ex vivo* with an expression vector directing the expression of a peptide having B7-2-like activity alone, or in conjunction with a peptide having B7-1-like activity and/or B7-3-like activity. The transfected tumor cells are returned to the patient to result in expression of the peptides on the surface of the transfected cell. Alternatively, gene therapy techniques can be used to target a tumor cell for transfection *in vivo*.

The presence of the peptide encoded by extended cDNAs derived from the 5' ESTs of the present invention having the activity of a B lymphocyte antigen(s) on the surface of the tumor cell provides the necessary costimulation signal to T cells to induce a T cell mediated immune response against the transfected tumor cells. In addition, tumor cells which lack or which fail to reexpress sufficient amounts of MHC class I or MHC class II molecules can be transfected with nucleic acids encoding all or a portion of (e.g., a cytoplasmic-domain truncated portion) of an MHC class I α chain and β_2 microglobulin or an MHC class II α chain and an MHC class II β chain to thereby express MHC class I or MHC class II proteins on the cell surface, respectively. Expression of the appropriate MHC class I or class II

molecules in conjunction with a peptide having the activity of a B lymphocyte antigen (e.g., B7-1, B7-2, B7-3) induces a T cell mediated immune response against the transfected tumor cell. Optionally, a gene encoding an antisense construct which blocks expression of an MHC class II associated protein, such as the invariant chain, can also be cotransfected with a DNA encoding a peptide having the activity of a B lymphocyte antigen to promote presentation of tumor associated antigens and induce tumor specific immunity. Thus, the induction of a T cell mediated immune response in a human subject may be sufficient to overcome tumor-specific tolerance in the subject. Alternatively, as described in more detail below, genes encoding these immune system regulator proteins or nucleic acids regulating the expression of such proteins may be introduced into appropriate host cells to increase or decrease the expression of the proteins as desired.

5

10

15

20

25

30

EXAMPLE 34

Assaying the Proteins Expressed from Extended cDNAs or Portions Thereof for Hematopoiesis Regulating Activity

The proteins encoded by the extended cDNAs or portions thereof may also be evaluated for their hematopoiesis regulating activity. For example, the effect of the proteins on embryonic stem cell differentiation may be evaluated. Numerous assays for such activity are familiar to those skilled in the art, including the assays disclosed in the following references, which are incorporated herein by reference: Johansson *et al. Cell. Biol.* 15:141-151, 1995; Keller *et al.*, *Mol. Cell. Biol.* 13:473-486, 1993; McClanahan *et al.*, *Blood* 81:2903-2915, 1993.

The proteins encoded by the extended cDNAs or portions thereof may also be evaluated for their influence on the lifetime of stem cells and stem cell differentiation. Numerous assays for such activity are familiar to those skilled in the art, including the assays disclosed in the following references, which are incorporated herein by reference: Freshney, Methylcellulose Colony Forming Assays, in *Culture of Hematopoietic Cells.*, Freshney, et al.. Eds. pp. 265-268, Wiley-Liss, Inc., New York, NY. 1994; Hirayama et al., Proc. Natl. Acad. Sci. USA 89:5907-5911, 1992; McNiece and Briddell, in Culture of Hematopoietic Cells, supra; Neben et al., Exp. Hematol. 22:353-359, 1994; Ploemacher and Cobblestone In

72

Culture of Hematopoietic Cells, supral-21, Spooncer et al, in Culture of Hematopoietic Cells, supral63-179 and Sutherland in Culture of Hematopoietic Cells, supra. 139-162.

Those proteins which exhibit hematopoiesis regulatory activity may then be formulated as pharmaceuticals and used to treat clinical conditions in which regulation of hematopoeisis is beneficial, such as in the treatment of myeloid or lymphoid cell deficiencies. Involvement in regulating hematopoiesis is indicated even by marginal biological activity in support of colony forming cells or of factor-dependent cell lines. For example, proteins supporting the growth and proliferation of erythroid progenitor cells alone or in combination with other cytokines, indicates utility, for example, in treating various anemias or for use in conjunction with irradiation/chemotherapy to stimulate the production of erythroid precursors and/or erythroid cells. Proteins supporting the growth and proliferation of myeloid cells such as granulocytes and monocytes/macrophages (i.e., traditional CSF activity) may be useful, for example, in conjunction with chemotherapy to prevent or treat consequent myelosuppression. Proteins supporting the growth and proliferation of megakaryocytes and consequently of platelets allows prevention or treatment of various platelet disorders such as thrombocytopenia, and generally may be used in place of or complementary to platelet transfusions. Proteins supporting the growth and proliferation of hematopoietic stem cells which are capable of maturing to any and all of the above-mentioned hematopoietic cells may therefore find therapeutic utility in various stem cell disorders (such as those usually treated with transplantion, including, without limitation, aplastic anemia and paroxysmal noctumal hemoglobinuria), as well as in repopulating the stem cell compartment post irradiation/chemotherapy, either in vivo or ex vivo (i.e., in conjunction with bone marrow transplantation or with peripheral progenitor cell transplantation (homologous or heterologous)) as normal cells or genetically manipulated for gene therapy. Alternatively, as described in more detail below, genes encoding hematopoiesis regulating activity proteins or nucleic acids regulating the expression of such proteins may be introduced into appropriate host cells to increase or decrease the expression of the proteins as desired.

5

10

15

20

25

73

EXAMPLE 35

Assaying the Proteins Expressed from Extended cDNAs or Portions Thereof for Regulation of Tissue Growth

The proteins encoded by the extended cDNAs or portions thereof may also be evaluated for their effect on tissue growth. Numerous assays for such activity are familiar to those skilled in the art, including the assays disclosed in International Patent Publication No. WO95/16035, International Patent Publication No. WO95/05846 and International Patent Publication No. WO91/07491, which are incorporated herein by reference.

5

10

15

20

25

30

Assays for wound healing activity include, without limitation, those described in: Winter, *Epidermal Wound Healing*, pps. 71-112, Maibach and Rovee, eds., Year Book Medical Publishers, Inc., Chicago, as modified by Eaglstein and Mertz, *J. Invest. Dermatol.* 71:382-84, 1978, which are incorporated herein by reference.

Those proteins which are involved in the regulation of tissue growth may then be formulated as pharmaceuticals and used to treat clinical conditions in which regulation of tissue growth is beneficial. For example, a protein encoded by extended cDNAs derived from the 5' ESTs of the present invention also may have utility in compositions used for bone, cartilage, tendon, ligament and/or nerve tissue growth or regeneration, as well as for wound healing and tissue repair and replacement, and in the treatment of burns, incisions and ulcers.

A protein encoded by extended cDNAs derived from the 5' ESTs of the present invention, which induces cartilage and/or bone growth in circumstances where bone is not normally formed, has application in the healing of bone fractures and cartilage damage or defects in humans and other animals. Such a preparation employing a protein of the invention may have prophylactic use in closed as well as open fracture reduction and also in the improved fixation of artificial joints. *De novo* bone synthesis induced by an osteogenic agent contributes to the repair of congenital, trauma induced, or oncologic resection induced craniofacial defects, and also is useful in cosmetic plastic surgery.

A protein of this invention may also be used in the treatment of periodontal disease, and in other tooth repair processes. Such agents may provide an environment to attract bone-forming cells, stimulate growth of bone-forming cells or induce differentiation of bone-forming cell progenitors. A protein of the invention may also be useful in the treatment of osteoporosis or osteoarthritis, such as through stimulation of bone and/or cartilage repair or

by blocking inflammation or processes of tissue destruction (collagenase activity, osteoclast activity, etc.) mediated by inflammatory processes.

5

10

15

20

25

30

Another category of tissue regeneration activity that may be attributable to the protein encoded by extended cDNAs derived from the 5' ESTs of the present invention is tendon/ligament formation. A protein encoded by extended cDNAs derived from the 5' ESTs of the present invention, which induces tendon/ligament-like tissue or other tissue formation in circumstances where such tissue is not normally formed, has application in the healing of tendon or ligament tears, deformities and other tendon or ligament defects in humans and other animals. Such a preparation employing a tendon/ligament-like tissue inducing protein may have prophylactic use in preventing damage to tendon or ligament tissue, as well as use in the improved fixation of tendon or ligament to bone or other tissues, and in repairing defects to tendon or ligament tissue. De novo tendon/ligament-like tissue formation induced by a composition encoded by extended cDNAs derived from the 5' ESTs of the present invention contributes to the repair of tendon or ligaments defects of congenital, traumatic or other origin and is also useful in cosmetic plastic surgery for attachment or repair of tendons or ligaments. The compositions encoded by extended cDNAs derived from the 5' ESTs of the present invention may provide an environment to attract tendon- or ligamentforming cells, stimulate growth of tendon- or ligament-forming cells, induce differentiation of progenitors of tendon- or ligament-forming cells, or induce growth of tendon/ligament cells or progenitors ex vivo for return in vivo to effect tissue repair. The compositions of the invention may also be useful in the treatment of tendinitis, carpal tunnel syndrome and other tendon or ligament defects. The compositions may also include an appropriate matrix and/or sequestering agent as a carrier as is well known in the art.

The protein encoded by extended cDNAs derived from the 5' ESTs of the present invention may also be useful for proliferation of neural cells and for regeneration of nerve and brain tissue, *i.e.*, for the treatment of central and peripheral nervous system diseases and neuropathies, as well as mechanical and traumatic disorders, which involve degeneration, death or trauma to neural cells or nerve tissue. More specifically, a protein may be used in the treatment of diseases of the peripheral nervous system, such as peripheral nerve injuries, peripheral neuropathy and localized neuropathies, and central nervous system diseases, such as Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic lateral sclerosis, and

75

Shy-Drager syndrome. Further conditions which may be treated in accordance with the present invention include mechanical and traumatic disorders, such as spinal cord disorders, head trauma and cerebrovascular diseases such as stroke. Peripheral neuropathies resulting from chemotherapy or other medical therapies may also be treatable using a protein of the invention.

Proteins of the invention may also be useful to promote better or faster closure of non-healing wounds, including without limitation pressure ulcers, ulcers associated with vascular insufficiency, surgical and traumatic wounds, and the like.

It is expected that a protein encoded by extended cDNAs derived from the 5' ESTs of the present invention may also exhibit activity for generation or regeneration of other tissues, such as organs (including, for example, pancreas, liver, intestine, kidney, skin, endothelium) muscle (smooth, skeletal or cardiac) and vascular (including vascular endothelium) tissue, or for promoting the growth of cells comprising such tissues. Part of the desired effects may be by inhibition or modulation of fibrotic scarring to allow normal tissue to generate. A protein of the invention may also exhibit angiogenic activity.

A protein encoded by extended cDNAs derived from the 5' ESTs of the present invention may also be useful for gut protection or regeneration and treatment of lung or liver fibrosis, reperfusion injury in various tissues, and conditions resulting from systemic cytokinc damage.

A protein encoded by extended cDNAs derived from the 5' ESTs of the present invention may also be useful for promoting or inhibiting differentiation of tissues described above from precursor tissues or cells, or for inhibiting the growth of tissues described above.

Alternatively, as described in more detail below, genes encoding tissue growth regulating activity proteins or nucleic acids regulating the expression of such proteins may be introduced into appropriate host cells to increase or decrease the expression of the proteins as desired.

25

5

10

15

20

76

EXAMPLE 36

Assaying the Proteins Expressed from Extended cDNAs or Portions Thereof for Regulation of Reproductive Hormones

5

10

15

20

25

30

The proteins encoded by the extended cDNAs or portions thereof may also be evaluated for their ability to regulate reproductive hormones, such as follicle stimulating hormone. Numerous assays for such activity are familiar to those skilled in the art, including the assays disclosed in the following references, which are incorporated herein by reference: Vale et al., Endocrinol. 91:562-572, 1972; Ling et al., Nature 321:779-782, 1986; Vale et al., Nature 321:776-779, 1986; Mason et al., Nature 318:659-663, 1985; Forage et al., Proc. Natl. Acad. Sci. USA 83:3091-3095, 1986, Chapter 6.12 in Current Protocols in Immunology, Coligan et al. Eds. Greene Publishing Associates and Wiley-Intersciece; Taub et al., J. Clin. Invest. 95:1370-1376, 1995; Lind et al., APMIS 103:140-146, 1995; Muller et al., Eur. J. Immunol. 25:1744-1748; Gruber et al., J. Immunol. 152:5860-5867, 1994; Johnston et al., J Immunol. 153:1762-1768, 1994.

Those proteins which exhibit activity as reproductive hormones or regulators of cell movement may then be formulated as pharmaceuticals and used to treat clinical conditions in which regulation of reproductive hormones are beneficial. For example, a protein encoded by extended cDNAs derived from the 5' ESTs of the present invention may also exhibit activinor inhibin-related activities. Inhibins are characterized by their ability to inhibit the release of follicle stimulating hormone (FSH), while activins are characterized by their ability to stimulate the release of FSH. Thus, a protein encoded by extended cDNAs derived from the 5' ESTs of the present invention, alone or in heterodimers with a member of the inhibin $\boldsymbol{\alpha}$ family, may be useful as a contraceptive based on the ability of inhibins to decrease fertility in female mammals and decrease spermatogenesis in male mammals. Administration of sufficient amounts of other inhibins can induce infertility in these mammals. Alternatively, the protein of the invention, as a homodimer or as a heterodimer with other protein subunits of the inhibin-B group, may be useful as a fertility inducing therapeutic, based upon the ability of activin molecules in stimulating FSH release from cells of the anterior pituitary. See, for example, United States Patent 4,798,885, the disclosure of which is incorporated herein by reference. A protein of the invention may also be useful for advancement of the onset of

77

fertility in sexually immature mammals, so as to increase the lifetime reproductive performance of domestic animals such as cows, sheep and pigs.

Alternatively, as described in more detail below, genes encoding reproductive hormone regulating activity proteins or nucleic acids regulating the expression of such proteins may be introduced into appropriate host cells to increase or decrease the expression of the proteins as desired.

EXAMPLE 37

Assaying the Proteins Expressed from Extended cDNAs or

10 Portions Thereof for Chemotactic/Chemokinetic Activity

5

15

20

25

30

The proteins encoded by the extended cDNAs or portions thereof may also be evaluated for chemotactic/chemokinetic activity. For example, a protein encoded by extended cDNAs derived from the 5' ESTs of the present invention may have chemotactic or chemokinetic activity (e.g., act as a chemokine) for mammalian cells, including, for example, monocytes, fibroblasts, neutrophils, T-cells, mast cells, eosinophils, epithelial and/or endothelial cells. Chemotactic and chemokinetic proteins can be used to mobilize or attract a desired cell population to a desired site of action. Chemotactic or chemokinetic proteins provide particular advantages in treatment of wounds and other trauma to tissues, as well as in treatment of localized infections. For example, attraction of lymphocytes, monocytes or neutrophils to tumors or sites of infection may result in improved immune responses against the tumor or infecting agent.

A protein or peptide has chemotactic activity for a particular cell population if it can stimulate, directly or indirectly, the directed orientation or movement of such cell population. Preferably, the protein or peptide has the ability to directly stimulate directed movement of cells. Whether a particular protein has chemotactic activity for a population of cells can be readily determined by employing such protein or peptide in any known assay for cell chemotaxis.

The activity of a protein of the invention may, among other means, be measured by the following methods:

Assays for chemotactic activity (which will identify proteins that induce or prevent chemotaxis) consist of assays that measure the ability of a protein to induce the migration of

cells across a membrane as well as the ability of a protein to induce the adhesion of one cell population to another cell population. Suitable assays for movement and adhesion include, without limitation, those described in: Current Protocols in Immunology, Ed by Coligan, Kruisbeek, Margulies, Shevach and Strober, Pub. Greene Publishing Associates and Wiley-Interscience, Chapter 6.12: 6.12.1-6.12.28; Taub et al., J. Clin. Invest. 95:1370-1376, 1995; Lind et al., APMIS 103:140-146, 1995; Mueller et al., Eur. J. Immunol. 25:1744-1748; Gruber et al., J. Immunol. 152:5860-5867, 1994; Johnston et al. J. Immunol., 153:1762-1768, 1994.

10

15

20

25

30

5

EXAMPLE 38

Assaying the Proteins Expressed from Extended cDNAs or Portions Thereof for Regulation of Blood Clotting

The proteins encoded by the extended cDNAs or portions thereof may also be evaluated for their effects on blood clotting. Numerous assays for such activity are familiar to those skilled in the art, including the assays disclosed in the following references, which are incorporated herein by reference: Linet et al., J. Clin. Pharmacol. 26:131-140, 1986; Burdick et al., Thrombosis Res. 45:413-419, 1987; Humphrey et al., Fibrinolysis 5:71-79, 1991; Schaub, Prostaglandins 35:467-474, 1988.

Those proteins which are involved in the regulation of blood clotting may then be formulated as pharmaceuticals and used to treat clinical conditions in which regulation of blood clotting is beneficial. For example, a protein of the invention may also exhibit hemostatic or thrombolytic activity. As a result, such a protein is expected to be useful in treatment of various coagulations disorders (including hereditary disorders, such as hemophilias) or to enhance coagulation and other hemostatic events in treating wounds resulting from trauma, surgery or other causes. A protein of the invention may also be useful for dissolving or inhibiting formation of thromboses and for treatment and prevention of conditions resulting therefrom (such as infarction of cardiac and central nervous system vessels (e.g., stroke)). Alternatively, as described in more detail below, genes encoding blood clotting activity proteins or nucleic acids regulating the expression of such proteins may be introduced into appropriate host cells to increase or decrease the expression of the proteins as desired.

79

EXAMPLE 39

Assaying the Proteins Expressed from Extended cDNAs or Portions Thereof for Involvement in Receptor/Ligand Interactions

The proteins encoded by the extended cDNAs or a portion thereof may also be evaluated for their involvement in receptor/ligand interactions. Numerous assays for such involvement are familiar to those skilled in the art, including the assays disclosed in the following references, which are incorporated herein by reference: Chapter 7. 7.28.1-7.28.22 in Current Protocols in Immunology, Coligan et al. Eds. Greene Publishing Associates and Wiley-Interscience; Takai et al., Proc. Natl. Acad. Sci. USA 84:6864-6868, 1987; Bierer et al., J. Exp. Med. 168:1145-1156, 1988; Rosenstein et al., J. Exp. Med. 169:149-160, 1989; Stoltenborg et al., J. Immunol. Methods 175:59-68, 1994; Stitt et al., Cell 80:661-670, 1995; Gyuris et al., Cell 75:791-803, 1993.

For example, the proteins encoded by extended cDNAs derived from the 5' ESTs of the present invention may also demonstrate activity as receptors, receptor ligands or inhibitors or agonists of receptor/ligand interactions. Examples of such receptors and ligands include, without limitation, cytokine receptors and their ligands, receptor kinases and their ligands, receptor phosphatases and their ligands, receptors involved in cell-cell interactions and their ligands (including without limitation, cellular adhesion molecules (such as selectins, integrins and their ligands) and receptor/ligand pairs involved in antigen presentation, antigen recognition and development of cellular and humoral immune responses). Receptors and ligands are also useful for screening of potential peptide or small molecule inhibitors of the relevant receptor/ligand interaction. A protein encoded by extended cDNAs derived from the 5' ESTs of the present invention (including, without limitation, fragments of receptors and ligands) may themselves be useful as inhibitors of receptor/ligand interactions. Alternatively, as described in more detail below, genes encoding proteins involved in receptor/ligand interactions or nucleic acids regulating the expression of such proteins may be introduced into appropriate host cells to increase or decrease the expression of the proteins as desired.

5

10

15

20

25

80

EXAMPLE 40

Assaying the Proteins Expressed from Extended cDNAs or Portions Thereof for Anti-Inflammatory Activity

5

10

15

20

The proteins encoded by the extended cDNAs or a portion thereof may also be evaluated for anti-inflammatory activity. The anti-inflammatory activity may be achieved by providing a stimulus to cells involved in the inflammatory response, by inhibiting or promoting cell-cell interactions (such as, for example, cell adhesion), by inhibiting or promoting chemotaxis of cells involved in the inflammatory process, inhibiting or promoting cell extravasation, or by stimulating or suppressing production of other factors which more directly inhibit or promote an inflammatory response. Proteins exhibiting such activities can be used to treat inflammatory conditions including chronic or acute conditions, including without limitation inflammation associated with infection (such as septic shock, sepsis or systemic inflammatory response syndrome), ischemia-reperfusioninury, endotoxin lethality, arthritis, complement-mediated hyperacute rejection, nephritis, cytokine- or chemokineinduced lung injury, inflammatory bowel disease, Crohn's disease or resulting from over production of cytokines such as TNF or IL-1. Proteins of the invention may also be useful to treat anaphylaxis and hypersensitivity to an antigenic substance or material. Alternatively, as described in more detail below, genes encoding anti-inflammatory activity proteins or nucleic acids regulating the expression of such proteins may be introduced into appropriate host cells to increase or decrease the expression of the proteins as desired.

EXAMPLE 41

Assaying the Proteins Expressed from Extended cDNAs or Portions Thereof for Tumor Inhibition Activity

The proteins encoded by the extended cDNAs or a portion thereof may also be evaluated for tumor inhibition activity. In addition to the activities described above for immunological treatment or prevention of tumors, a protein of the invention may exhibit other anti-tumor activities. A protein may inhibit tumor growth directly or indirectly (such as, for example, via ADCC). A protein may exhibit its tumor inhibitory activity by acting on tumor tissue or tumor precursor tissue, by inhibiting formation of tissues necessary to support tumor growth (such as, for example, by inhibiting angiogenesis), by causing production of other

81

factors, agents or cell types which inhibit tumor growth, or by suppressing, eliminating or inhibiting factors, agents or cell types which promote tumor growth. Alternatively, as described in more detail below, genes tumor inhibition activity proteins or nucleic acids regulating the expression of such proteins may be introduced into appropriate host cells to increase or decrease the expression of the proteins as desired.

A protein of the invention may also exhibit one or more of the following additional activities or effects: inhibiting the growth, infection or function of, or killing, infectious agents, including, without limitation, bacteria, viruses, fungi and other parasites; effecting (suppressing or enhancing) bodily characteristics, including, without limitation, height, weight, hair color, eye color, skin, fat to lean ratio or other tissue pigmentation, or organ or body part size or shape (such as, for example, breast augmentation or diminution, change in bone form or shape); effecting biorhythms or circadian cycles or rhythms; effecting the fertility of male or female subjects; effecting the metabolism, catabolism, anabolism, processing, utilization, storage or elimination of dietary fat, lipid, protein, carbohydrate, vitamins, minerals, cofactors or other nutritional factors or component(s); effecting behavioral characteristics, including, without limitation, appetite, libido, stress, cognition (including cognitive disorders), depression (including depressive disorders) and violent behaviors; providing analgesic effects or other pain reducing effects; promoting differentiation and growth of embryonic stem cells in lineages other than hematopoietic lineages; hormonal or endocrine activity; in the case of enzymes, correcting deficiencies of the enzyme and treating deficiency-related diseases; treatment of hyperproliferative disorders (such as, for example, psoriasis); immunoglobulin-like activity (such as, for example, the ability to bind antigens or complement); and the ability to act as an antigen in a vaccine composition to raise an immune response against such protein or another material or entity which is cross-reactive with such protein. Alternatively, as described in more detail below, genes encoding proteins involved in any of the above mentioned activities or nucleic acids regulating the expression of such proteins may be introduced into appropriate host cells to increase or decrease the expression of the proteins as desired.

5

10

15

20

25

82

EXAMPLE 42

Identification of Proteins which Interact with Polypeptides Encoded by Extended cDNAs

5

10

15

20

25

30

Proteins which interact with the polypeptides encoded by cDNAs derived from the 5' ESTs or fragments thereof, such as receptor proteins, may be identified using two hybrid systems such as the Matchmaker Two Hybrid System 2 (Catalog No. K1604-1, Clontech). As described in the manual accompanying the kit which is incorporated herein by reference, the the cDNAs derived from 5' ESTs, or fragments thereof, are inserted into an expression vector such that they are in frame with DNA encoding the DNA binding domain of the yeast transcriptional activator GALA. cDNAs in a cDNA library which encode proteins which might interact with the polypeptides encoded by the extended cDNAs or portions thereof are inserted into a second expression vector such that they are in frame with DNA encoding the activation domain of GAL4. The two expression plasmids are transformed into yeast and the yeast are plated on selection medium which selects for expression of selectable markers on each of the expression vectors as well as GALA dependent expression of the HIS3 gene. Transformants capable of growing on medium lacking histidine are screened for GAL4 dependent lacZ expression. Those cells which are positive in both the histidine selection and the lacZ assay contain plasmids encoding proteins which interact with the polypeptide encoded by the extended cDNAs or portions thereof.

Alternatively, the system described in Lustig et al., Methods in Enzymology 283: 83-99, 1997, and in U.S. Patent No. 5,654,150, the disclosure of which is incorporated herein by reference, may be used for identifying molecules which interact with the polypeptides encoded by extended cDNAs. In such systems, in vitro transcription reactions are performed on a pool of vectors containing extended cDNA inserts cloned downstream of a promoter which drives in vitro transcription. The resulting pools of mRNAs are introduced into Xenopus laevis oocytes. The oocytes are then assayed for a desired activity.

Alternatively, the pooled *in vitro* transcription products produced as described above may be translated *in vitro*. The pooled *in vitro* translation products can be assayed for a desired activity or for interaction with a known polypeptide.

Proteins or other molecules interacting with polypeptides encoded by extended cDNAs can be found by a variety of additional techniques. In one method, affinity columns containing the polypeptide encoded by the extended cDNA or a portion thereof

83

can be constructed. In some versions, of this method the affinity column contains chimeric proteins in which the protein encoded by the extended cDNA or a portion thereof is fused to glutathione S-transferase. A mixture of cellular proteins or pool of expressed proteins as described above and is applied to the affinity column. Proteins interacting with the polypeptide attached to the column can then be isolated and analyzed on 2-D electrophoresis gel as described in Ramunsen et al., Electrophoresis 18:588-598, 1997, the disclosure of which is incorporated herein by reference. Alternatively, the proteins retained on the affinity column can be purified by electrophoresis based methods and sequenced. The same method can be used to isolate antibodies, to screen phage display products, or to screen phage display human antibodies.

5

10

15

20

25

30

Proteins interacting with polypeptides encoded by extended cDNAs or portions thereof can also be screened by using an Optical Biosensor as described in Edwards and Leatherbarrow, Analytical Biochemistry 246:1-6, 1997, the disclosure of which is incorporated herein by reference. The main advantage of the method is that it allows the determination of the association rate between the protein and other interacting molecules. Thus, it is possible to specifically select interacting molecules with a high or low association rate. Typically a target molecule is linked to the sensor surface (through a carboxymethl dextran matrix) and a sample of test molecules is placed in contact with the target molecules. The binding of a test molecule to the target molecule causes a change in the refractive index and/ or thickness. This change is detected by the Biosensor provided it occurs in the evanescent field (which extend a few hundred nanometers from the sensor surface). In these screening assays, the target molecule can be one of the polypeptides encoded by extended cDNAs or a portion thereof and the test sample can be a collection of proteins extracted from tissues or cells, a pool of expressed proteins, combinatorial peptide and/ or chemical libraries, or phage displayed peptides. The tissues or cells from which the test proteins are extracted can originate from any species.

In other methods, a target protein is immobilized and the test population is a collection of unique polypeptides encoded by the extended cDNAs or portions thereof.

To study the interaction of the proteins encoded by the extended cDNAs or portions thereof with drugs, the microdialysis coupled to HPLC method described by

84

Wang et al., Chromatographia 44:205-208, 1997 or the affinity capillary electrophoresis method described by Busch et al., J. Chromatogr. 777:311-328, 1997, the disclosures of which are incorporated herein by reference can be used.

It will be appreciated by those skilled in the art that the proteins expressed from the extended cDNAs or portions may be assayed for numerous activities in addition to those specifically enumerated above. For example, the expressed proteins may be evaluated for applications involving control and regulation of inflammation, tumor proliferation or metastasis, infection, or other clinical conditions. In addition, the proteins expressed from the extended cDNAs or portions thereof may be useful as nutritional agents or cosmetic agents.

The proteins expressed from the cDNAs or portions thereof may be used to generate antibodies capable of specifically binding to the expressed protein or fragments thereof as described in Example 40 below. The antibodies may capable of binding a full length protein encoded by a cDNA derived from a 5' EST, a mature protein (*i.e.* the protein generated by cleavage of the signal peptide) encoded by a cDNA derived from a 5' EST, or a signal peptide encoded by a cDNA derived from a 5' EST. Alternatively, the antibodies may be capable of binding fragments of at least 10 amino acids of the proteins encoded by the above cDNAs. In some embodiments, the antibodies may be capable of binding fragments of at least 15 amino acids of the proteins encoded by the above cDNAs. In other embodiments, the antibodies may be capable of binding fragments of at least 25 amino acids of the proteins expressed from the extended cDNAs which comprise at least 25 amino acids of the proteins encoded by the above cDNAs. In further embodiments, the antibodies may be capable of binding fragments of at least 25 amino acids of the proteins

25

30

5

10

15

20

EXAMPLE 43

Production of an Antibody to a Human Protein

Substantially pure protein or polypeptide is isolated from the transfected or transformed cells as described in Example 30. The concentration of protein in the final preparation is adjusted, for example, by concentration on an Amicon filter device, to the level of a few µg/ml. Monoclonal or polyclonal antibody to the protein can then be prepared as follows:

85

1. Monoclonal Antibody Production by Hybridoma Fusion

Monoclonal antibody to epitopes of any of the peptides identified and isolated as described can be prepared from murine hybridomas according to the classical method of Kohler, and Milstein, Nature 256:495, 1975 or derivative methods thereof. Briefly, a mouse is repetitively inoculated with a few micrograms of the selected protein or peptides derived therefrom over a period of a few weeks. The mouse is then sacrificed, and the antibody producing cells of the spleen isolated. The spleen cells are fused by means of polyethylene glycol with mouse myeloma cells, and the excess unfused cells destroyed by growth of the system on selective media comprising aminopterin (HAT media). The successfully fused cells are diluted and aliquots of the dilution placed in wells of a microtiter plate where growth of the culture is continued. Antibody-producing clones are identified by detection of antibody in the supernatant fluid of the wells by immunoassay procedures, such as ELISA, as originally described by Engvall, Meth. Enzymol. 70:419, 1980, the disclosure of which is incorporated herein by reference and derivative methods thereof. Selected positive clones can be expanded and their monoclonal antibody product harvested for use. Detailed procedures for monoclonal antibody production are described in Davis et al. in Basic Methods in Molecular Biology Elsevier, New York. Section 21-2, the disclosure of which is incorporated herein by reference.

20

25

30

5

10

15

2. Polyclonal Antibody Production by Immunization

Polyclonal antiserum containing antibodies to heterogenous epitopes of a single protein can be prepared by immunizing suitable animals with the expressed protein or peptides derived therefrom, which can be unmodified or modified to enhance immunogenicity. Effective polyclonal antibody production is affected by many factors related both to the antigen and the host species. For example, small molecules tend to be less immunogenic than others and may require the use of carriers and adjuvant. Also, host animals response vary depending on site of inoculations and doses, with both inadequate or excessive doses of antigen resulting in low titer antisera. Small doses (ng level) of antigen administered at multiple intradermal sites appears to be most reliable. An effective

86

immunization protocol for rabbits can be found in Vaitukaitis. et al, J. Clin. Endocrinol. Metab. 33:988-991 (1971), the disclosure of which is incorporated herein by reference.

Booster injections can be given at regular intervals, and antiserum harvested when antibody titer thereof, as determined semi-quantitatively, for example, by double immunodiffusion in agar against known concentrations of the antigen, begins to fall. See, for example, Ouchterlony, *et al.*, Chap. 19 in: *Handbook of Experimental Immunology* D. Wier (ed) Blackwell (1973), the disclosure of which is incorporated herein by reference. Plateau concentration of antibody is usually in the range of 0.1 to 0.2 mg/ml of serum (about 12 μM). Affinity of the antisera for the antigen is determined by preparing competitive binding curves, as described, for example, by Fisher, D., Chap. 42 in: *Manual of Clinical Immunology*, 2d Ed. (Rose and Friedman, Eds.) Amer. Soc. For Microbiol., Washington, D.C. (1980), the disclosure of which is incorporated herein by reference.

Antibody preparations prepared according to either protocol are useful in quantitative immunoassays which determine concentrations of antigen-bearing substances in biological samples; they are also used semi-quantitatively or qualitatively to identify the presence of antigen in a biological sample. The antibodies may also be used in therapeutic compositions for killing cells expressing the protein or reducing the levels of the protein in the body.

V. Use of 5' ESTs or Sequences Obtainable Therefrom or Portions Thereof as Reagents

The 5' ESTs of the present invention (or cDNAs or genomic DNAs obtainable therefrom) may be used as reagents in isolation procedures, diagnostic assays, and forensic procedures. For example, sequences from the 5' ESTs (or cDNAs or genomic DNAs obtainable therefrom) may be detectably labeled and used as probes to isolate other sequences capable of hybridizing to them. In addition, sequences from 5' ESTs (or cDNAs or genomic DNAs obtainable therefrom) may be used to design PCR primers to be used in isolation, diagnostic, or forensic procedures.

25

5

10

15

87

1. Use of 5' ESTs or Sequences Obtainable Therefrom or Portions Thereof in Isolation,

Diagnostic and Forensic Procedures

EXAMPLE 44

5

10

15

20

Preparation of PCR Primers and Amplification of DNA

The 5' EST sequences (or cDNAs or genomic DNAs obtainable therefrom) may be used to prepare PCR primers for a variety of applications, including isolation procedures for cloning nucleic acids capable of hybridizing to such sequences, diagnostic techniques and forensic techniques. The PCR primers are at least 10 bases, and preferably at least 12, 15, or 17 bases in length. More preferably, the PCR primers are at least 20-30 bases in length. In some embodiments, the PCR primers may be more than 30 bases in length. It is preferred that the primer pairs have approximately the same G/C ratio, so that melting temperatures are approximately the same. A variety of PCR techniques are familiar to those skilled in the art. For a review of PCR technology, see Molecular Cloning to Genetic Engineering, White Ed. in Methods in Molecular Biology 67: Humana Press, Totowa 1997, the disclosure of which is incorporated herein by reference. In each of these PCR procedures, PCR primers on either side of the nucleic acid sequences to be amplified are added to a suitably prepared nucleic acid sample along with dNTPs and a thermostable polymerase such as Taq polymerase, Pfu polymerase, or Vent polymerase. The nucleic acid in the sample is denatured and the PCR primers are specifically hybridized to complementary nucleic acid sequences in the sample. The hybridized primers are extended. Thereafter, another cycle of denaturation, hybridization, and extension is initiated. The cycles are repeated multiple times to produce an amplified fragment containing the nucleic acid sequence between the primer sites.

25

30

EXAMPLE 45

Use of 5'ESTs as Probes

Probes derived from 5' ESTs (or cDNAs or genomic DNAs obtainable therefrom), including full length cDNAs or genomic sequences, may be labeled with detectable labels familiar to those skilled in the art, including radioisotopes and non-radioactive labels, to provide a detectable probe. The detectable probe may be single stranded or double stranded and may be made using techniques known in the art, including *in vitro* transcription, nick

translation, or kinase reactions. A nucleic acid sample containing a sequence capable of hybridizing to the labeled probe is contacted with the labeled probe. If the nucleic acid in the sample is double stranded, it may be denatured prior to contacting the probe. In some applications, the nucleic acid sample may be immobilized on a surface such as a nitrocellulose or nylon membrane. The nucleic acid sample may comprise nucleic acids obtained from a variety of sources, including genomic DNA, cDNA libraries, RNA, or tissue samples.

Procedures used to detect the presence of nucleic acids capable of hybridizing to the detectable probe include well known techniques such as Southern blotting, Northern blotting, dot blotting, colony hybridization, and plaque hybridization. In some applications, the nucleic acid capable of hybridizing to the labeled probe may be cloned into vectors such as expression vectors, sequencing vectors, or *in vitro* transcription vectors to facilitate the characterization and expression of the hybridizing nucleic acids in the sample. For example, such techniques may be used to isolate and clone sequences in a genomic library or cDNA library which are capable of hybridizing to the detectable probe as described in Example 30 above.

PCR primers made as described in Example 44 above may be used in forensic analyses, such as the DNA fingerprinting techniques described in Examples 46-50 below. Such analyses may utilize detectable probes or primers based on the sequences of the the 5' ESTs or of cDNAs or genomic DNAs isolated using the 5' ESTs.

20

25

30

5

10

15

EXAMPLE 46

Forensic Matching by DNA Sequencing

In one exemplary method, DNA samples are isolated from forensic specimens of, for example, hair, semen, blood or skin cells by conventional methods. A panel of PCR primers based on a number of the 5' ESTs of Example 25, or cDNAs or genomic DNAs isolated therefrom as described above, is then utilized in accordance with Example 44 to amplify DNA of approximately 100-200 bases in length from the forensic specimen. Corresponding sequences are obtained from a test subject. Each of these identification DNAs is then sequenced using standard techniques, and a simple database comparison determines the differences, if any, between the sequences from the subject and those from the sample. Statistically significant differences between the suspect's DNA sequences and those from the sample conclusively prove a lack of identity. This lack of identity can be proven, for example,

89

with only one sequence. Identity, on the other hand, should be demonstrated with a large number of sequences, all matching. Preferably, a minimum of 50 statistically identical sequences of 100 bases in length are used to prove identity between the suspect and the sample.

5

10

15

EXAMPLE 47

Positive Identification by DNA Sequencing

The technique outlined in the previous example may also be used on a larger scale to provide a unique fingerprint-type identification of any individual. In this technique, primers are prepared from a large number of 5'EST sequences from Example 25, or cDNA or genomic DNA sequences obtainable therefrom. Preferably, 20 to 50 different primers are used. These primers are used to obtain a corresponding number of PCR-generated DNA segments from the individual in question in accordance with Example 44. Each of these DNA segments is sequenced, using the methods set forth in Example 46. The database of sequences generated through this procedure uniquely identifies the individual from whom the sequences were obtained. The same panel of primers may then be used at any later time to absolutely correlate tissue or other biological specimen with that individual.

EXAMPLE 48

20

25

30

Southern Blot Forensic Identification

The procedure of Example 47 is repeated to obtain a panel of at least 10 amplified sequences from an individual and a specimen. Preferably, the panel contains at least 50 amplified sequences. More preferably, the panel contains 100 amplified sequences. In some embodiments, the panel contains 200 amplified sequences. This PCR-generated DNA is then digested with one or a combination of, preferably, four base specific restriction enzymes. Such enzymes are commercially available and known to those of skill in the art. After digestion, the resultant gene fragments are size separated in multiple duplicate wells on an agarose gel and transferred to nitrocellulose using Southern blotting techniques well known to those with skill in the art. For a review of Southern blotting see Davis *et al.* (Basic Methods in Molecular Biology, 1986, Elsevier Press. pp 62-65), the disclosure of which is incorporated herein by reference.

A panel of probes based on the sequences of 5' ESTs (or cDNAs or genomic DNAs obtainable therefrom), or fragments thereof of at least 10 bases, are radioactively or colorimetrically labeled using methods known in the art, such as nick translation or end labeling, and hybridized to the Southern blot using techniques known in the art (Davis *et al.*, supra). Preferably, the probe comprises at least 12, 15, or 17 consecutive nucleotides from the 5' EST (or cDNAs or genomic DNAs obtainable therefrom). More preferably, the probe comprises at least 20-30 consecutive nucleotides from the 5' EST (or cDNAs or genomic DNAs obtainable therefrom). In some embodiments, the probe comprises more than 30 nucleotides from the 5' EST (or cDNAs or genomic DNAs obtainable therefrom).

Preferably, at least 5 to 10 of these labeled probes are used, and more preferably at least about 20 or 30 are used to provide a unique pattern. The resultant bands appearing from the hybridization of a large sample of 5' EST (or cDNAs or genomic DNAs obtainable therefrom) will be a unique identifier. Since the restriction enzyme cleavage will be different for every individual, the band pattern on the Southern blot will also be unique. Increasing the number of 5' EST (or cDNAs or genomic DNAs obtainable therefrom) probes will provide a statistically higher level of confidence in the identification since there will be an increased number of sets of bands used for identification.

EXAMPLE 49

20

25

30

5

10

15

Dot Blot Identification Procedure

Another technique for identifying individuals using the 5' EST sequences disclosed herein utilizes a dot blot hybridization technique.

Genomic DNA is isolated from nuclei of subject to be identified. Oligonucleotide probes of approximately 30 bp in length are synthesized that correspond to at least 10, preferably 50 sequences from the 5' ESTs or cDNAs or genomic DNAs obtainable therefrom. The probes are used to hybridize to the genomic DNA through conditions known to those in the art. The oligonucleotides are end labeled with P³² using polynucleotide kinase (Pharmacia). Dot Blots are created by spotting the genomic DNA onto nitrocellulose or the like using a vacuum dot blot manifold (BioRad, Richmond California). The nitrocellulose filter containing the genomic sequences is baked or UV linked to the filter, prehybridized and hybridized with labeled probe using techniques known in the art (Davis et al., supra). The ³²P

labeled DNA fragments are sequentially hybridized with successively stringent conditions to detect minimal differences between the 30 bp sequence and the DNA. Tetramethylammonium chloride is useful for identifying clones containing small numbers of nucleotide mismatches (Wood et al., Proc. Natl. Acad. Sci. USA 82(6):1585-1588, 1985) which is hereby incorporated by reference. A unique pattern of dots distinguishes one individual from another individual.

5' EST sequences (or cDNAs or genomic DNAs obtainable therefrom) or oligonucleotides containing at least 10 consecutive bases from these sequences can be used as probes in the following alternative fingerprinting technique. Preferably, the probe comprises at least 12, 15, or 17 consecutive nucleotides from the 5' EST sequences (or cDNAs or genomic DNAs obtainable therefrom). More preferably, the probe comprises at least 20-30 consecutive nucleotides from the 5' EST sequences (or cDNAs or genomic DNAs obtainable therefrom). In some embodiments, the probe comprises more than 30 nucleotides from the 5' EST sequences (or cDNAs or genomic DNAs obtainable therefrom).

Preferably, a plurality of probes having sequences from different genes are used in the alternative fingerprinting technique. Example 50 below provides a representative alternative fingerprinting procedure in which the probes are derived from 5'EST.

EXAMPLE 50

20

25

5

10

15

Alternative "Fingerprint" Identification Technique

20-mer oligonucleotides are prepared from a large number, e.g. 50, 100, or 200, of 5'EST using commercially available oligonucleotide services such as Genset, Paris, France. Cell samples from the test subject are processed for DNA using techniques well known to those with skill in the art. The nucleic acid is digested with restriction enzymes such as EcoRI and XbaI. Following digestion, samples are applied to wells for electrophoresis. The procedure, as known in the art, may be modified to accommodate polyacrylamide electrophoresis, however in this example, samples containing 5 ug of DNA are loaded into wells and separated on 0.8% agarose gels. The gels are transferred onto nitrocellulose using standard Southern blotting techniques.

30 10 ng of each of the oligonucleotides are pooled and end-labeled with ³²P. The nitrocellulose is prehybridized with blocking solution and hybridized with the labeled probes.

Following hybridization and washing, the nitrocellulose filter is exposed to X-Omat AR X-ray film. The resulting hybridization pattern will be unique for each individual.

It is additionally contemplated within this example that the number of probe sequences used can be varied for additional accuracy or clarity.

5

The proteins encoded by the extended cDNAs may also be used to generate antibodies as explained in Examples 30 and 43 in order to identify the tissue type or cell species from which a sample is derived as described in example 51.

10

15

20

25

30

EXAMPLE 51

Identification of Tissue Types or Cell Species by Means of Labeled Tissue Specific Antibodies

Identification of specific tissues is accomplished by the visualization of tissue specific antigens by means of antibody preparations according to Examples 30 and 43 which are conjugated, directly or indirectly to a detectable marker. Selected labeled antibody species bind to their specific antigen binding partner in tissue sections, cell suspensions, or in extracts of soluble proteins from a tissue sample to provide a pattern for qualitative or semi-qualitative interpretation.

Antisera for these procedures must have a potency exceeding that of the native preparation, and for that reason, antibodies are concentrated to a mg/ml level by isolation of the gamma globulin fraction, for example, by ion-exchange chromatography or by ammonium sulfate fractionation. Also, to provide the most specific antisera, unwanted antibodies, for example to common proteins, must be removed from the gamma globulin fraction, for example by means of insoluble immunoabsorbents, before the antibodies are labeled with the marker. Either monoclonal or heterologous antisera is suitable for either procedure.

A. Immunohistochemical techniques

Purified, high-titer antibodies, prepared as described above, are conjugated to a detectable marker, as described, for example, by Fudenberg, Chap. 26 in: Basic and Clinical Immunology, 3rd Ed. Lange, Los Altos, California, 1980, or Rose, et al., Chap. 12 in:

Methods in Immunodiagnosis, 2d Ed. John Wiley and Sons, New York (1980), the disclosures of which are incorporated herein by reference.

A fluorescent marker, either fluorescein or rhodamine, is preferred, but antibodies can also be labeled with an enzyme that supports a color producing reaction with a substrate, such as horseradish peroxidase. Markers can be added to tissue-bound antibody in a second step, as described below. Alternatively, the specific antitissue antibodies can be labeled with ferritin or other electron dense particles, and localization of the ferritin coupled antigen-antibody complexes achieved by means of an electron microscope. In yet another approach, the antibodies are radiolabeled, with, for example ¹²⁵I, and detected by overlaying the antibody treated preparation with photographic emulsion

5

10

15

20

25

Preparations to carry out the procedures can comprise monoclonal or polyclonal antibodies to a single protein or peptide identified as specific to a tissue type, for example, brain tissue, or antibody preparations to several antigenically distinct tissue specific antigens can be used in panels, independently or in mixtures, as required.

Tissue sections and cell suspensions are prepared for immunohistochemical examination according to common histological techniques. Multiple cryostat sections (about 4 µm, unfixed) of the unknown tissue and known control, are mounted and each slide covered with different dilutions of the antibody preparation. Sections of known and unknown tissues should also be treated with preparations to provide a positive control, a negative control, for example, pre-immune sera, and a control for non-specific staining, for example, buffer.

Treated sections are incubated in a humid chamber for 30 min at room temperature, rinsed, then washed in buffer for 30-45 min. Excess fluid is blotted away, and the marker developed.

If the tissue specific antibody was not labeled in the first incubation, it can be labeled at this time in a second antibody-antibody reaction, for example, by adding fluorescein- or enzyme-conjugated antibody against the immunoglobulin class of the antiserum-producing species, for example, fluorescein labeled antibody to mouse IgG. Such labeled sera are commercially available.

The antigen found in the tissues by the above procedure can be quantified by measuring the intensity of color or fluorescence on the tissue section, and calibrating that signal using appropriate standards.

B. Identification of tissue specific soluble proteins

5

10

15

20

25

30

The visualization of tissue specific proteins and identification of unknown tissues from that procedure is carried out using the labeled antibody reagents and detection strategy as described for immunohistochemistry; however the sample is prepared according to an electrophoretic technique to distribute the proteins extracted from the tissue in an orderly array on the basis of molecular weight for detection.

A tissue sample is homogenized using a Virtis apparatus; cell suspensions are disrupted by Dounce homogenization or osmotic lysis, using detergents in either case as required to disrupt cell membranes, as is the practice in the art. Insoluble cell components such as nuclei, microsomes, and membrane fragments are removed by ultracentrifugation, and the soluble protein-containing fraction concentrated if necessary and reserved for analysis.

A sample of the soluble protein solution is resolved into individual protein species by conventional SDS polyacrylamide electrophoresis as described, for example, by Davis, et al., Section 19-2 in: Basic Methods in Molecular Biology, Leder ed., Elsevier, New York, 1986, the disclosure of which is incorporated herein by reference, using a range of amounts of polyacrylamide in a set of gels to resolve the entire molecular weight range of proteins to be detected in the sample. A size marker is run in parallel for purposes of estimating molecular weights of the constituent proteins. Sample size for analysis is a convenient volume of from 5 to 55 μ l, and containing from about 1 to 100 μ g protein. An aliquot of each of the resolved proteins is transferred by blotting to a nitrocellulose filter paper, a process that maintains the pattern of resolution. Multiple copies are prepared. The procedure, known as Western Blot Analysis, is well described in Davis, L. et al., supra Section 19-3. One set of nitrocellulose blots is stained with Coomassie blue dye to visualize the entire set of proteins for comparison with the antibody bound proteins. The remaining nitrocellulose filters are then incubated with a solution of one or more specific antisera to tissue specific proteins prepared as described in Examples 30 and 43. In this procedure, as in procedure A above, appropriate positive and negative sample and reagent controls are run.

In either procedure A or B, a detectable label can be attached to the primary tissue antigen-primary antibody complex according to various strategies and permutations thereof. In a straightforward approach, the primary specific antibody can be labeled; alternatively, the unlabeled complex can be bound by a labeled secondary anti-IgG antibody. In other approaches, either the primary or secondary antibody is conjugated to a biotin molecule, which can, in a subsequent step, bind an avidin conjugated marker. According to yet another strategy, enzyme labeled or radioactive protein A, which has the property of binding to any IgG, is bound in a final step to either the primary or secondary antibody.

5

10

15

20

25

30

The visualization of tissue specific antigen binding at levels above those seen in control tissues to one or more tissue specific antibodies, prepared from the gene sequences identified from extended cDNA sequences, can identify tissues of unknown origin, for example, forensic samples, or differentiated tumor tissue that has metastasized to foreign bodily sites.

In addition to their applications in forensics and identification, 5' ESTs (or cDNAs or genomic DNAs obtainable therefrom) may be mapped to their chromosomal locations. Example 52 below describes radiation hybrid (RH) mapping of human chromosomal regions using 5'ESTs. Example 53 below describes a representative procedure for mapping an 5' EST to its location on a human chromosome. Example 54 below describes mapping of 5' ESTs on metaphase chromosomes by Fluorescence In Situ Hybridization (FISH). Those skilled in the art will appreciate that the method of Examples 52-54 may also be used to map cDNAs or genomic DNAs obtainable from the 5' ESTs to their chromosomal locations.

2. Use of 5' ESTs or Sequences Obtainable Therefrom or Portions Thereof in Chromosome Mapping

EXAMPLE 52

Radiation hybrid mapping of 5'ESTs to the human genome

Radiation hybrid (RH) mapping is a somatic cell genetic approach that can be used for high resolution mapping of the human genome. In this approach, cell lines containing one or more human chromosomes are lethally irradiated, breaking each chromosome into fragments whose size depends on the radiation dose. These fragments are rescued by fusion

5

10

15

25

30

with cultured rodent cells, yielding subclones containing different portions of the human genome. This technique is described by Benham et al., Genomics 4:509-517, 1989; and Cox et al., Science 250:245-250, 1990, the entire contents of which are hereby incorporated by reference. The random and independent nature of the subclones permits efficient mapping of any human genome marker. Human DNA isolated from a panel of 80-100 cell lines provides a mapping reagent for ordering 5'EST. In this approach, the frequency of breakage between markers is used to measure distance, allowing construction of fine resolution maps as has been done using conventional ESTs (Schuler et al., Science 274:540-546, 1996, hereby incorporated by reference).

RH mapping has been used to generate a high-resolution whole genome radiation hybrid map of human chromosome 17q22-q25.3 across the genes for growth hormone (GH) and thymidine kinase (TK) (Foster et al., Genomics 33:185-192, 1996), the region surrounding the Gorlin syndrome gene (Obermayr et al., Eur. J. Hum. Genet. 4:242-245, 1996), 60 loci covering the entire short arm of chromosome 12 (Raeymaekers et al., Genomics 29:170-178, 1995), the region of human chromosome 22 containing the neurofibromatosis type 2 locus (Frazer et al., Genomics 14:574-584, 1992) and 13 loci on the long arm of chromosome 5 (Warrington et al., Genomics 11:701-708, 1991).

EXAMPLE 53

20 <u>Mapping of 5'ESTs to HumanChromosomes using PCR techniques</u>

5' ESTs (or cDNAs or genomic DNAs obtainable therefrom) may be assigned to human chromosomes using PCR based methodologies. In such approaches, oligonucleotide primer pairs are designed from the 5' ESTs (or cDNAs or genomic DNAs obtainable therefrom) to minimize the chance of amplifying through an intron. Preferably, the oligonucleotide primers are 18-23 bp in length and are designed for PCR amplification. The creation of PCR primers from known sequences is well known to those with skill in the art. For a review of PCR technology see Erlich in PCR Technology, Principles and Applications for DNA Amplification, Freeman and Co., New York, 1992, the disclosure of which is incorporated herein by reference.

The primers are used in polymerase chain reactions (PCR) to amplify templates from total human genomic DNA. PCR conditions are as follows: 60 ng of genomic DNA is used

5

10

15

20

as a template for PCR with 80 ng of each oligonucleotide primer, 0.6 unit of Taq polymerase, and 1 μCu of a ³²P-labeled deoxycytidine triphosphate. The PCR is performed in a microplate thermocycler (Techne) under the following conditions: 30 cycles of 94°C, 1.4 min; 55°C, 2 min; and 72°C, 2 min; with a final extension at 72°C for 10 min. The amplified products are analyzed on a 6% polyacrylamide sequencing gel and visualized by autoradiography. If the length of the resulting PCR product is identical to the distance between the ends of the primer sequences in the extended cDNA from which the primers are derived, then the PCR reaction is repeated with DNA templates from two panels of human-rodent somatic cell hybrids, BIOS PCRable DNA (BIOS Corporation) and NIGMS Human-Rodent Somatic Cell Hybrid Mapping Panel Number 1 (NIGMS, Camden, NJ).

PCR is used to screen a series of somatic cell hybrid cell lines containing defined sets of human chromosomes for the presence of a given 5' EST (or cDNA or genomic DNA obtainable therefrom). DNA is isolated from the somatic hybrids and used as starting templates for PCR_reactions using the primer pairs from the 5' EST (or cDNA or genomic DNA obtainable therefrom). Only those somatic cell hybrids with chromosomes containing the human gene corresponding to the 5' EST (or cDNA or genomic DNA obtainable therefrom) will yield an amplified fragment. The 5' EST (or cDNA or genomic DNA obtainable therefrom) are assigned to a chromosome by analysis of the segregation pattern of PCR products from the somatic hybrid DNA templates. The single human chromosome present in all cell hybrids that give rise to an amplified fragment is the chromosome containing that 5'EST (or cDNA or genomic DNA obtainable therefrom). For a review of techniques and analysis of results from somatic cell gene mapping experiments, see Ledbetter *et al.*, *Genomics* 6:475-481, 1990, the disclosure of which is incorporated herein by reference.

25

30

EXAMPLE 54

Mapping of Extended 5' ESTs to Chromosomes Using Fluorescence In Situe Hybridization

Fluorescence in situ hybridization allows the 5'EST (or cDNA or genomic DNA obtainable therefrom) to be mapped to a particular location on a given chromosome. The chromosomes to be used for fluorescence in situ hybridization techniques may be obtained from a variety of sources including cell cultures, tissues, or whole blood.

In a preferred embodiment, chromosomal localization of an 5'EST (or cDNA or genomic DNA obtainable therefrom) is obtained by FISH as described by Cherif et al. (Proc. Natl. Acad. Sci. U.S.A., 87:6639-6643, 1990), the disclosure of which is incorporated herein by reference.. Metaphase chromosomes are prepared from phytohemagglutinin (PHA)stimulated blood cell donors. PHA-stimulated lymphocytes from healthy males are cultured for 72 h in RPMI-1640 medium. For synchronization, methotrexate (10 μM) is added for 17 h, followed by addition of 5-bromodeoxyuridine (5-BrdU, 0.1 mM) for 6 h. Colcemid (1 $\mu g/ml$) is added for the last 15 min before harvesting the cells. Cells are collected, washed in RPMI, incubated with a hypotonic solution of KCl (75 mM) at 37°C for 15 min and fixed in three changes of methanol:acetic acid (3:1). The cell suspension is dropped onto a glass slide and air dried. The 5'EST (or cDNA or genomic DNA obtainable therefrom) is labeled with biotin-16 dUTP by nick translation according to the manufacturer's instructions (Bethesda Research Laboratories, Bethesda, MD), purified using a Sephadex G-50 column (Pharmacia, Upsala, Sweden) and precipitated. Just prior to hybridization, the DNA pellet is dissolved in hybridization buffer (50% formamide, 2 X SSC, 10% dextran sulfate, 1 mg/ml sonicated salmon sperm DNA, pH 7) and the probe is denatured at 70°C for 5-10 min.

5

10

15

20

25

30

Slides kept at -20°C are treated for 1 h at 37°C with RNase A (100 µg/ml), rinsed three times in 2 X SSC and dehydrated in an ethanol series. Chromosome preparations are denatured in 70% formamide, 2 X SSC for 2 min at 70°C, then dehydrated at 4°C. The slides are treated with proteinase K (10 µg/100 ml in 20 mM Tris-HCl, 2 mM CaCl₂) at 37°C for 8 min and dehydrated. The hybridization mixture containing the probe is placed on the slide, covered with a coverslip, sealed with rubber cement and incubated overnight in a humid chamber at 37°C. After hybridization and post-hybridization washes, the biotinylated probe is detected by avidin-FITC and amplified with additional layers of biotinylated goat anti-avidin and avidin-FITC. For chromosomal localization, fluorescent R-bands are obtained as previously described (Cherif et al., supra.). The slides are observed under a LEICA fluorescence microscope (DMRXA). Chromosomes are counterstained with propidium iodide and the fluorescent signal of the probe appears as two symmetrical yellow-green spots on both chromatids of the fluorescent R-band chromosome (red). Thus, a particular 5'EST (or cDNA or genomic DNA obtainable therefrom) may be localized to a particular cytogenetic R-band on a given chromosome.

99

Once the 5'EST (or cDNA or genomic DNA obtainable therefrom) have been assigned to particular chromosomes using the techniques described in Examples 52-54 above, they may be utilized to construct a high resolution map of the chromosomes on which they are located or to identify the chromosomes in a sample.

5

10

15

20

25

EXAMPLE 55

Use of 5'EST to Construct or Expand Chromosome Maps

Chromosome mapping involves assigning a given unique sequence to a particular chromosome as described above. Once the unique sequence has been mapped to a given chromosome, it is ordered relative to other unique sequences located on the same chromosome. One approach to chromosome mapping utilizes a series of yeast artificial chromosomes (YACs) bearing several thousand long inserts derived from the chromosomes of the organism from which the extended cDNAs (or genomic DNAs obtainable therefrom) are obtained. This approach is described in Nagaraja et al., Genome Research 7:210-222, 1997, the disclosure of which is incorporated herein by reference. Briefly, in this approach each chromosome is broken into overlapping pieces which are inserted into the YAC vector. The YAC inserts are screened using PCR or other methods to determine whether they include the 5'EST (or cDNA or genomic DNA obtainable therefrom) whose position is to be determined. Once an insert has been found which includes the 5'EST (or cDNA or genomic DNA obtainable therefrom), the insert can be analyzed by PCR or other methods to determine whether the insert also contains other sequences known to be on the chromosome or in the region from which the 5'EST (or cDNA or genomic DNA obtainable therefrom) was derived. This process can be repeated for each insert in the YAC library to determine the location of each of the extended cDNAs (or genomic DNAs obtainable therefrom) relative to one another and to other known chromosomal markers. In this way, a high resolution map of the distribution of numerous unique markers along each of the organisms chromosomes may be obtained.

As described in Example 56 below extended cDNAs (or genomic DNAs obtainable therefrom) may also be used to identify genes associated with a particular phenotype, such as hereditary disease or drug response.

100

3. Use of 5'ESTs or Sequences Obtained Therefrom or Fragments Thereof in Gene Identification

EXAMPLE 56

Identification of genes associated with hereditary diseases or drug response

This example illustrates an approach useful for the association of 5'ESTs (or cDNA or genomic DNA obtainable therefrom) with particular phenotypic characteristics. In this example, a particular 5'EST (or cDNA or genomic DNA obtainable therefrom) is used as a test probe to associate that 5'EST (or cDNA or genomic DNA obtainable therefrom) with a particular phenotypic characteristic.

5'ESTs (or cDNA or genomic DNA obtainable therefrom) are mapped to a particular location on a human chromosome using techniques such as those described in Examples 52 and 53 or other techniques known in the art. A search of Mendelian Inheritance in Man (McKusick in *Mendelian Inheritance in Man* (available on line through Johns Hopkins University Welch Medical Library) reveals the region of the human chromosome which contains the 5'EST (or cDNA or genomic DNA obtainable therefrom) to be a very gene rich region containing several known genes and several diseases or phenotypes for which genes have not been identified. The gene corresponding to this 5'EST (or cDNA or genomic DNA obtainable therefrom) thus becomes an immediate candidate for each of these genetic diseases.

Cells from patients with these diseases or phenotypes are isolated and expanded in culture. PCR primers from the 5'EST (or cDNA or genomic DNA obtainable therefrom) are used to screen genomic DNA, mRNA or cDNA obtained from the patients. 5'ESTs (or cDNA or genomic DNA obtainable therefrom) that are not amplified in the patients can be positively associated with a particular disease by further analysis. Alternatively, the PCR analysis may yield fragments of different lengths when the samples are derived from an individual having the phenotype associated with the disease than when the sample is derived from a healthy individual, indicating that the gene containing the 5'EST may be responsible for the genetic disease.

5

10

15

20

25

VI. Use of 5'EST (or cDNA or Genomic DNA Obtainable Therefrom) to Construct Vectors

The present 5'ESTs (or cDNA or genomic DNA obtainable therefrom) may also be used to construct secretion vectors capable of directing the secretion of the proteins encoded by genes therein. Such secretion vectors may facilitate the purification or enrichment of the proteins encoded by genes inserted therein by reducing the number of background proteins from which the desired protein must be purified or enriched. Exemplary secretion vectors are described in Example 57 below.

10 <u>1. Construction of Secretion Vectors</u>

5

15

20

25

30

EXAMPLE 57

Construction of Secretion Vectors

The secretion vectors include a promoter capable of directing gene expression in the host cell, tissue, or organism of interest. Such promoters include the Rous Sarcoma Virus promoter, the SV40 promoter, the human cytomegalovirus promoter, and other promoters familiar to those skilled in the art.

A signal sequence from a 5' EST (or cDNAs or genomic DNAs obtainable therefrom) is operably linked to the promoter such that the mRNA transcribed from the promoter will direct the translation of the signal peptide. The host cell, tissue, or organism may be any cell, tissue, or organism which recognizes the signal peptide encoded by the signal sequence in the 5' EST (or cDNA or genomic DNA obtainable therefrom). Suitable hosts include mammalian cells, tissues or organisms, avian cells, tissues, or organisms, insect cells, tissues or organisms, or yeast.

In addition, the secretion vector contains cloning sites for inserting genes encoding the proteins which are to be secreted. The cloning sites facilitate the cloning of the insert gene in frame with the signal sequence such that a fusion protein in which the signal peptide is fused to the protein encoded by the inserted gene is expressed from the mRNA transcribed from the promoter. The signal peptide directs the extracellular secretion of the fusion protein.

The secretion vector may be DNA or RNA and may integrate into the chromosome of the host, be stably maintained as an extrachromosomal replicon in the host, be an artificial chromosome, or be transiently present in the host. Many nucleic acid backbones suitable for

use as secretion vectors are known to those skilled in the art, including retroviral vectors, SV40 vectors, Bovine Papilloma Virus vectors, yeast integrating plasmids, yeast episomal plasmids, yeast artificial chromosomes, human artificial chromosomes, P element vectors, baculovirus vectors, or bacterial plasmids capable of being transiently introduced into the host.

The secretion vector may also contain a polyA signal such that the polyA signal is located downstream of the gene inserted into the secretion vector.

5

10

15

20

25

30

After the gene encoding the protein for which secretion is desired is inserted into the secretion vector, the secretion vector is introduced into the host cell, tissue, or organism using calcium phosphate precipitation, DEAE-Dextran, electroporation, liposome-mediated transfection, viral particles or as naked DNA. The protein encoded by the inserted gene is then purified or enriched from the supernatant using conventional techniques such as ammonium sulfate precipitation, immunoprecipitation, immunochromatography, size exclusion chromatography, ion exchange chromatography, and HPLC. Alternatively, the secreted protein may be in a sufficiently enriched or pure state in the supernatant or growth media of the host to permit it to be used for its intended purpose without further enrichment.

The signal sequences may also be inserted into vectors designed for gene therapy. In such vectors, the signal sequence is operably linked to a promoter such that mRNA transcribed from the promoter encodes the signal peptide. A cloning site is located downstream of the signal sequence such that a gene encoding a protein whose secretion is desired may readily be inserted into the vector and fused to the signal sequence. The vector is introduced into an appropriate host cell. The protein expressed from the promoter is secreted extracellularly, thereby producing a therapeutic effect.

The 5' ESTs may also be used to clone sequences located upstream of the 5' ESTs which are capable of regulating gene expression, including promoter sequences, enhancer sequences, and other upstream sequences which influence transcription or translation levels. Once identified and cloned, these upstream regulatory sequences may be used in expression vectors designed to direct the expression of an inserted gene in a desired spatial, temporal, developmental, or quantitative fashion. Example 58 describes a method for cloning sequences upstream of the extended cDNAs or 5' ESTs.

103

2. Identification of Upstream Sequences With Promoting or Regulatory Activities EXAMPLE 58

5

10

15

20

25

30

Use of Extended cDNAs or 5' ESTs to Clone Upstream Sequences from Genomic DNA

Sequences derived from extended cDNAs or 5' ESTs may be used to isolate the promoters of the corresponding genes using chromosome walking techniques. In one chromosome walking technique, which utilizes the GenomeWalkerTM kit available from Clontech, five complete genomic DNA samples are each digested with a different restriction enzyme which has a 6 base recognition site and leaves a blunt end. Following digestion, oligonucleotide adapters are ligated to each end of the resulting genomic DNA fragments.

For each of the five genomic DNA libraries, a first PCR reaction is performed according to the manufacturer's instructions (which are incorporated herein by reference) using an outer adaptor primer provided in the kit and an outer gene specific primer. The gene specific primer should be selected to be specific for the extended cDNA or 5' EST of interest and should have a melting temperature, length, and location in the extended cDNA or 5'EST which is consistent with its use in PCR reactions. Each first PCR reaction contains 5 ng of genomic DNA, 5 µl of 10X Tth reaction buffer, 0.2 mM of each dNTP, 0.2 µM each of outer adaptor primer and outer gene specific primer, 1.1 mM of Mg(OAc)₂, and 1 µl of the Tth polymerase 50X mix in a total volume of 50 µl. The reaction cycle for the first PCR reaction is as follows: 1 min - 94°C / 2 sec - 94°C, 3 min - 72°C (7 cycles) / 2 sec - 94°C, 3 min - 67°C.

The product of the first PCR reaction is diluted and used as a template for a second PCR reaction according to the manufacturer's instructions using a pair of nested primers which are located internally on the amplicon resulting from the first PCR reaction. For example, 5 µl of the reaction product of the first PCR reaction mixture may be diluted 180 times. Reactions are made in a 50 µl volume having a composition identical to that of the first PCR reaction except the nested primers are used. The first nested primer is specific for the adaptor, and is provided with the GenomeWalkerTM kit. The second nested primer is specific for the particular extended cDNA or 5' EST for which the promoter is to be cloned and should have a melting temperature, length, and location in the extended cDNA or 5' EST which is consistent with its use in PCR reactions. The reaction parameters of the second PCR reaction are as follows: 1 min -

104

94°C / 2 sec - 94°C, 3 min - 72°C (6 cycles) / 2 sec - 94°C, 3 min - 67°C (25 cycles) / 5 min - 67°C. The product of the second PCR reaction is purified, cloned, and sequenced using standard techniques.

Alternatively, two or more human genomic DNA libraries can be constructed by using two or more restriction enzymes. The digested genomic DNA is cloned into vectors which can be converted into single stranded, circular, or linear DNA. A biotinylated oligonucleotide comprising at least 15 nucleotides from the extended cDNA or 5' EST sequence is hybridized to the single stranded DNA. Hybrids between the biotinylated oligonucleotide and the single stranded DNA containing the extended cDNA or EST sequence are isolated as described in Example 29 above. Thereafter, the single stranded DNA containing the extended cDNA or EST sequence is released from the beads and converted into double stranded DNA using a primer specific for the extended cDNA or 5' EST sequence or a primer corresponding to a sequence included in the cloning vector. The resulting double stranded DNA is transformed into bacteria. DNAs containing the 5' EST or extended cDNA sequences are identified by colony PCR or colony hybridization.

Once the upstream genomic sequences have been cloned and sequenced as described above, prospective promoters and transcription start sites within the upstream sequences may be identified by comparing the sequences upstream of the extended cDNAs or 5' ESTs with databases containing known transcription start sites, transcription factor binding sites, or promoter sequences.

In addition, promoters in the upstream sequences may be identified using promoter reporter vectors as described in Example .

25 EXAMPLE 59

5

10

15

20

30

Identification of Promoters in Cloned Upstream Sequences

The genomic sequences upstream of the extended cDNAs or 5' ESTs are cloned into a suitable promoter reporter vector, such as the pSEAP-Basic, pSEAP-Enhancer, pβgal-Basic, pβgal-Enhancer, or pEGFP-1 Promoter Reporter vectors available from Clontech Briefly, each of these promoter reporter vectors include multiple cloning sites positioned upstream of a reporter gene encoding a readily assayable protein such as secreted alkaline

phosphatase, β galactosidase, or green fluorescent protein. The sequences upstream of the extended cDNAs or 5' ESTs are inserted into the cloning sites upstream of the reporter gene in both orientations and introduced into an appropriate host cell. The level of reporter protein is assayed and compared to the level obtained from a vector which lacks an insert in the cloning site. The presence of an elevated expression level in the vector containing the insert with respect to the control vector indicates the presence of a promoter in the insert. If necessary, the upstream sequences can be cloned into vectors which contain an enhancer for augmenting transcription levels from weak promoter sequences. A significant level of expression above that observed with the vector lacking an insert indicates that a promoter sequence is present in the inserted upstream sequence.

5

10

15

20

25

Appropriate host cells for the promoter reporter vectors may be chosen based on the results of the above described determination of expression patterns of the extended cDNAs and ESTs. For example, if the expression pattern analysis indicates that the mRNA corresponding to a particular extended cDNA or 5' EST is expressed in fibroblasts, the promoter reporter vector may be introduced into a human fibroblast cell line.

Promoter sequences within the upstream genomic DNA may be further defined by constructing nested deletions in the upstream DNA using conventional techniques such as Exonuclease III digestion. The resulting deletion fragments can be inserted into the promoter reporter vector to determine whether the deletion has reduced or obliterated promoter activity. In this way, the boundaries of the promoters may be defined. If desired, potential individual regulatory sites within the promoter may be identified using site directed mutagenesis or linker scanning to obliterate potential transcription factor binding sites within the promoter individually or in combination. The effects of these mutations on transcription levels may be determined by inserting the mutations into the cloning sites in the promoter reporter vectors.

EXAMPLE 60

Cloning and Identification of Promoters

Using the method described in Example 58 above with 5' ESTs, sequences upstream of several genes were obtained. Using the primer pairs GGG AAG ATG GAG ATA GTA TTG CCT G (SEQ ID NO:29) and CTG CCA TGT ACA TGA TAG AGA GAT TC (SEQ

106

ID NO:30), the promoter having the internal designation P13H2 (SEQ ID NO:31) was obtained.

Using the primer pairs GTA CCA GGGG ACT GTG ACC ATT GC (SEQ ID NO:32) and CTG TGA CCA TTG CTC CCA AGA GAG (SEQ ID NO:33), the promoter having the internal designation P15B4 (SEQ ID NO:34) was obtained.

5

10

15

20

25

30

Using the primer pairs CTG GGA TGG AAG GCA CGG TA (SEQ ID NO:35) and GAG ACC ACA CAG CTA GAC AA (SEQ ID NO:36), the promoter having the internal designation P29B6 (SEQ ID NO:37) was obtained.

Figure 4 provides a schematic description of the promoters isolated and the way they are assembled with the corresponding 5' tags. The upstream sequences were screened for the presence of motifs resembling transcription factor binding sites or known transcription start sites using the computer program MatInspector release 2.0, August 1996.

Table VII describes the transcription factor binding sites present in each of these promoters. The columns labeled matrice provides the name of the MatInspector matrix used. The column labeled position provides the 5' position of the promoter site. Numeration of the sequence starts from the transcription site as determined by matching the genomic sequence with the 5' EST sequence. The column labeled "orientation" indicates the DNA strand on which the site is found, with the + strand being the coding strand as determined by matching the genomic sequence with the sequence of the 5' EST. The column labeled "score" provides the MatInspector score found for this site. The column labeled "length" provides the length of the site in nucleotides. The column labeled "sequence" provides the sequence of the site found.

Bacterial clones containing plasmids containing the promoter sequences described above described above are presently stored in the inventor's laboratories under the internal identification numbers provided above. The inserts may be recovered from the deposited materials by growing an aliquot of the appropriate bacterial clone in the appropriate medium. The plasmid DNA can then be isolated using plasmid isolation procedures familiar to those skilled in the art such as alkaline lysis minipreps or large scale alkaline lysis plasmid isolation procedures. If desired the plasmid DNA may be further enriched by centrifugation on a cesium chloride gradient, size exclusion chromatography, or anion exchange chromatography. The plasmid DNA obtained using these procedures may then be manipulated using standard

107

cloning teehniques familiar to those skilled in the art. Alternatively, a PCR can be done with primers designed at both ends of the EST insertion. The PCR product which corresponds to the 5' EST can then be manipulated using standard cloning techniques familiar to those skilled in the art.

The promoters and other regulatory sequences located upstream of the extended cDNAs or 5' ESTs may be used to design expression vectors capable of directing the expression of an inserted gene in a desired spatial, temporal, developmental, or quantitative manner. A promoter capable of directing the desired spatial, temporal, developmental, and quantitative patterns may be selected using the results of the expression analysis described in Example 26 above. For example, if a promoter which confers a high level of expression in muscle is desired, the promoter sequence upstream of an extended cDNA or 5' EST derived from an mRNA which is expressed at a high level in muscle, as determined by the method of Example 26, may be used in the expression vector.

Preferably, the desired promoter is placed near multiple restriction sites to facilitate the cloning of the desired insert downstream of the promoter, such that the promoter is able to drive expression of the inserted gene. The promoter may be inserted in conventional nucleic acid backbones designed for extrachromosomal replication, integration into the host chromosomes or transient expression. Suitable backbones for the present expression vectors include retroviral backbones, backbones from eukaryotic episomes such as SV40 or Bovine Papilloma Virus, backbones from bacterial episomes, or artificial chromosomes.

Preferably, the expression vectors also include a polyA signal downstream of the multiple restriction sites for directing the polyadenylation of mRNA transcribed from the gene inserted into the expression vector.

Following the identification of promoter sequences using the procedures of Examples 58-60, proteins which interact with the promoter may be identified as described in Example 61 below.

5

10

15

EXAMPLE 61

Identification of Proteins Which Interact with Promoter Sequences, Upstream Regulatory Sequences, or mRNA

5

10

15

20

25

30

Sequences within the promoter region which are likely to bind transcription factors may be identified by homology to known transcription factor binding sites or through conventional mutagenesis or deletion analyses of reporter plasmids containing the promoter sequence. For example, deletions may be made in a reporter plasmid containing the promoter sequence of interest operably linked to an assayable reporter gene. The reporter plasmids carrying various deletions within the promoter region are transfected into an appropriate host cell and the effects of the deletions on expression levels is assessed. Transcription factor binding sites within the regions in which deletions reduce expression levels may be further localized using site directed mutagenesis, linker scanning analysis, or other techniques familiar to those skilled in the art.

Nucleic acids encoding proteins which interact with sequences in the promoter may be identified using one-hybrid systems such as those described in the manual accompanying the Matchmaker One-Hybrid System kit available from Clontech (Catalog No. K1603-1), the disclosure of which is incorporated herein by reference. Briefly, the Matchmaker One-hybrid system is used as follows. The target sequence for which it is desired to identify binding proteins is cloned upstream of a selectable reporter gene and integrated into the yeast genome. Preferably, multiple copies of the target sequences are inserted into the reporter plasmid in tandem. A library comprised of fusions between cDNAs to be evaluated for the ability to bind to the promoter and the activation domain of a yeast transcription factor, such as GAL4, is transformed into the yeast strain containing the integrated reporter sequence. The yeast are plated on selective media to select cells expressing the selectable marker linked to the promoter sequence. The colonies which grow on the selective media contain genes encoding proteins which bind the target sequence. The inserts in the genes encoding the fusion proteins are further characterized by sequencing. In addition, the inserts may be inserted into expression vectors or in vitro transcription vectors. Binding of the polypeptides encoded by the inserts to the promoter DNA may be confirmed by techniques familiar to those skilled in the art, such as gel shift analysis or DNAse protection analysis.

109

VII. Use of 5' ESTs (or cDNAs or Genomic DNAs Obtainable Therefrom) in Gene Therapy

The present invention also comprises the use of 5'ESTs (or cDNA or genomic DNA obtainable therefrom) in gene therapy strategies, including antisense and triple helix strategies as described in Examples 62 and 63 below. In antisense approaches, nucleic acid sequences complementary to an mRNA are hybridized to the mRNA intracellularly, thereby blocking the expression of the protein encoded by the mRNA. The antisense sequences may prevent gene expression through a variety of mechanisms. For example, the antisense sequences may inhibit the ability of ribosomes to translate the mRNA. Alternatively, the antisense sequences may block transport of the mRNA from the nucleus to the cytoplasm, thereby limiting the amount of mRNA available for translation. Another mechanism through which antisense sequences may inhibit gene expression is by interfering with mRNA splicing. In yet another strategy, the antisense nucleic acid may be incorporated in a ribozyme capable of specifically cleaving the target mRNA.

15

20

25

30

10

5

EXAMPLE 62

Preparation and Use of Antisense Oligonucleotides

The antisense nucleic acid molecules to be used in gene therapy may be either DNA or RNA sequences. They may comprise a sequence complementary to the sequence of the 5'EST (or cDNA or genomic DNA obtainable therefrom). The antisense nucleic acids should have a length and melting temperature sufficient to permit formation of an intracellular duplex with sufficient stability to inhibit the expression of the mRNA in the duplex. Strategies for designing antisense nucleic acids suitable for use in gene therapy are disclosed in Green et al., Ann. Rev. Biochem. 55:569-597, 1986; and Izant and Weintraub, Cell 36:1007-1015, 1984, which are hereby incorporated by reference.

In some strategies, antisense molecules are obtained from a nucleotide sequence encoding a protein by reversing the orientation of the coding region with respect to a promoter so as to transcribe the opposite strand from that which is normally transcribed in the cell. The antisense molecules may be transcribed using *in vitro* transcription systems such as those which employ T7 or SP6 polymerase to generate the transcript. Another approach

involves transcription of the antisense nucleic acids in vivo by operably linking DNA containing the antisense sequence to a promoter in an expression vector.

Alternatively, oligonucleotides which are complementary to the strand normally transcribed in the cell may be synthesized *in vitro*. Thus, the antisense nucleic acids are complementary to the corresponding mRNA and are capable of hybridizing to the mRNA to create a duplex. In some embodiments, the antisense sequences may contain modified sugar phosphate backbones to increase stability and make them less sensitive to RNase activity. Examples of modifications suitable for use in antisense strategies are described by Rossi *et al.*, *Pharmacol. Ther.* 50(2):245-254, 1991, which is hereby incorporated by reference.

5

10

15

20

25

30

Various types of antisense oligonucleotides complementary to the sequence of the 5'EST (or cDNA or genomic DNA obtainable therefrom) may be used. In one preferred embodiment, stable and semi-stable antisense oligonucleotides described in International Application No. PCT WO94/23026, hereby incorporated by reference, are used. In these molecules, the 3' end or both the 3' and 5' ends are engaged in intramolecular hydrogen bonding between complementary base pairs. These molecules are better able to withstand exonuclease attacks and exhibit increased stability compared to conventional antisense oligonucleotides.

In another preferred embodiment, the antisense oligodeoxynucleotides against herpes simplex virus types 1 and 2 described in International Application No. WO 95/04141, hereby incorporated by reference, are used.

In yet another preferred embodiment, the covalently cross-linked antisense oligonucleotides described in International Application No. WO 96/31523, hereby incorporated by reference, are used. These double- or single-stranded oligonucleotides comprise one or more, respectively, inter- or intra-oligonucleotide covalent cross-linkages, wherein the linkage consists of an amide bond between a primary amine group of one strand and a carboxyl group of the other strand or of the same strand, respectively, the primary amine group being directly substituted in the 2' position of the strand nucleotide monosaccharide ring, and the carboxyl group being carried by an aliphatic spacer group substituted on a nucleotide or nucleotide analog of the other strand or the same strand, respectively.

111

The antisense oligodeoxynucleotides and oligonucleotides disclosed in International Application No. WO 92/18522, incorporated by reference, may also be used. These molecules are stable to degradation and contain at least one transcription control recognition sequence which binds to control proteins and are effective as decoys therefore. These molecules may contain "hairpin" structures, "dumbbell" structures, "modified dumbbell" structures, "cross-linked" decoy structures and "loop" structures.

5

10

15

20

25

30

In another preferred embodiment, the cyclic double-stranded oligonucleotides described in European Patent Application No. 0 572 287 A2, hereby incorporated by reference are used. These ligated oligonucleotide "dumbbells" contain the binding site for a transcription factor and inhibit expression of the gene under control of the transcription factor by sequestering the factor.

Use of the closed antisense oligonucleotides disclosed in International Application No. WO 92/19732, hereby incorporated by reference, is also contemplated. Because these molecules have no free ends, they are more resistant to degradation by exonucleases than are conventional oligonucleotides. These oligonucleotides may be multifunctional, interacting with several regions which are not adjacent to the target mRNA.

The appropriate level of antisense nucleic acids required to inhibit gene expression may be determined using *in vitro* expression analysis. The antisense molecule may be introduced into the cells by diffusion, injection, infection, transfection or h-region-mediated import using procedures known in the art. For example, the antisense nucleic acids can be introduced into the body as a bare or naked oligonucleotide, oligonucleotide encapsulated in lipid, oligonucleotide sequence encapsidated by viral protein, or as an oligonucleotide operably linked to a promoter contained in an expression vector. The expression vector may be any of a variety of expression vectors known in the art, including retroviral or viral vectors, vectors capable of extrachromosomal replication, or integrating vectors. The vectors may be DNA or RNA.

The antisense molecules are introduced onto cell samples at a number of different concentrations preferably between $1\times10^{-10}M$ to $1\times10^{-4}M$. Once the minimum concentration that can adequately control gene expression is identified, the optimized dose is translated into a dosage suitable for use *in vivo*. For example, an inhibiting concentration in culture of 1×10^{-7} translates into a dose of approximately 0.6 mg/kg bodyweight. Levels of oligonucleotide

approaching 100 mg/kg bodyweight or higher may be possible after testing the toxicity of the oligonucleotide in laboratory animals. It is additionally contemplated that cells from the vertebrate are removed, treated with the antisense oligonucleotide, and reintroduced into the vertebrate.

It is further contemplated that the antisense oligonucleotide sequence is incorporated into a ribozyme sequence to enable the antisense to specifically bind and cleave its target mRNA. For technical applications of ribozyme and antisense oligonucleotides see Rossi et al., supra.

In a preferred application of this invention, the polypeptide encoded by the gene is first identified, so that the effectiveness of antisense inhibition on translation can be monitored using techniques that include but are not limited to antibody-mediated tests such as RIAs and ELISA, functional assays, or radiolabeling.

The 5' ESTs of the present invention (or cDNAs or genomic DNAs obtainable therefrom) may also be used in gene therapy approaches based on intracellular triple helix formation. Triple helix oligonucleotides are used to inhibit transcription from a genome. They are particularly useful for studying alterations in cell activity as it is associated with a particular gene. The 5' EST sequences (or cDNAs or genomic DNAs obtainable therefrom) of the present invention or, more preferably, a portion of those sequences, can be used to inhibit gene expression in individuals having diseases associated with expression of a particular gene. Similarly, a portion of 5' EST sequences (or cDNAs or genomic DNAs obtainable therefrom) can be used to study the effect of inhibiting transcription of a particular gene within a cell. Traditionally, homopurine sequences were considered the most useful for triple helix strategies. However, homopyrimidine sequences can also inhibit gene expression. Such homopyrimidine oligonucleotides bind to the major homopurine:homopyrimidine sequences. Thus, both types of sequences from the 5'EST or from the gene corresponding to the 5'EST are contemplated within the scope of this invention.

EXAMPLE 63

5

10

15

20

The sequences of the 5' ESTs (or cDNAs or genomic DNAs obtainable therefrom) are scanned to identify 10-mer to 20-mer homopyrimidine or homopurine stretches which could be used in triple-helix based strategies for inhibiting gene expression. Following identification of candidate homopyrimidine or homopurine stretches, their efficiency in inhibiting gene expression is assessed by introducing varying amounts of oligonucleotides containing the candidate sequences into tissue culture cells which normally express the target gene. The oligonucleotides may be prepared on an oligonucleotide synthesizer or they may be purchased commercially from a company specializing in custom oligonucleotide synthesis, such as GENSET, Paris, France.

5

10

15

20

25

30

The oligonucleotides may be introduced into the cells using a variety of methods known to those skilled in the art, including but not limited to calcium phosphate precipitation, DEAE-Dextran, electroporation, liposome-mediated transfection or native uptake.

Treated cells are monitored for altered cell function or reduced gene expression using techniques such as Northern blotting, RNase protection assays, or PCR based strategies to monitor the transcription levels of the target gene in cells which have been treated with the oligonucleotide. The cell functions to be monitored are predicted based upon the homologies of the target gene corresponding to the extended cDNA from which the oligonucleotide was derived with known gene sequences that have been associated with a particular function. The cell functions can also be predicted based on the presence of abnormal physiologies within cells derived from individuals with a particular inherited disease, particularly when the extended cDNA is associated with the disease using techniques described in Example 56.

The oligonucleotides which are effective in inhibiting gene expression in tissue culture cells may then be introduced *in vivo* using the techniques described above and in Example 62 at a dosage calculated based on the *in vitro* results, as described in Example 62.

In some embodiments, the natural (beta) anomers of the oligonucleotide units can be replaced with alpha anomers to render the oligonucleotide more resistant to nucleases. Further, an intercalating agent such as ethidium bromide, or the like, can be attached to the 3' end of the alpha oligonucleotide to stabilize the triple helix. For information on the generation of oligonucleotides suitable for triple helix formation see Griffin *et al.*, *Science* **245**:967-971, 1989, which is hereby incorporated by this reference.

114

EXAMPLE 64

Use of cDNAs Obtained Using the 5' ESTs to Express an Encoded Protein in a Host Organism

The cDNAs obtained as described above using the 5' ESTs of the present invention may also be used to express an encoded protein in a host organism to produce a beneficial effect. In such procedures, the encoded protein may be transiently expressed in the host organism or stably expressed in the host organism. The encoded protein may have any of the activities described above. The encoded protein may be a protein which the host organism lacks or, alternatively, the encoded protein may augment the existing levels of the protein in the host organism.

A full length extended cDNA encoding the signal peptide and the mature protein, or an extended cDNA encoding only the mature protein is introduced into the host organism. The extended cDNA may be introduced into the host organism using a variety of techniques known to those of skill in the art. For example, the extended cDNA may be injected into the host organism as naked DNA such that the encoded protein is expressed in the host organism, thereby producing a beneficial effect.

Alternatively, the extended cDNA may be cloned into an expression vector downstream of a promoter which is active in the host organism. The expression vector may be any of the expression vectors designed for use in gene therapy, including viral or retroviral vectors. The expression vector may be directly introduced into the host organism such that the encoded protein is expressed in the host organism to produce a beneficial effect. In another approach, the expression vector may be introduced into cells *in vitro*. Cells containing the expression vector are thereafter selected and introduced into the host organism, where they express the encoded protein to produce a beneficial effect.

25

30

5

10

15

20

EXAMPLE 65

<u>Use of Signal Peptides Encoded by 5' ESTs or Sequences obtained Therefrom</u> <u>to Import Proteins Into Cells</u>

The short core hydrophobic region (h) of signal peptides encoded by the 5'ESTS or extended cDNAs derived from SEQ ID NOs: 38-184 may also be used as a carrier to import a peptide or a protein of interest, so-called cargo, into tissue culture cells (Lin et al., J. Biol.

5

10

15

20

25

30

Chem., 270: 14225-14258, 1995; Du et al., J. Peptide Res., 51: 235-243, 1998; Rojas et al., Nature Biotech., 16: 370-375, 1998).

When cell permeable peptides of limited size (approximately up to 25 amino acids) are to be translocated across cell membrane, chemical synthesis may be used in order to add the h region to either the C-terminus or the N-terminus to the cargo peptide of interest. Alternatively, when longer peptides or proteins are to be imported into cells, nucleic acids can be genetically engineered, using techniques familiar to those skilled in the art, in order to link the extended cDNA sequence encoding the h region to the 5' or the 3' end of a DNA sequence coding for a cargo polypeptide. Such genetically engineered nucleic acids are then translated either *in vitro* or *in vivo* after transfection into appropriate cells, using conventional techniques to produce the resulting cell permeable polypeptide. Suitable hosts cells are then simply incubated with the cell permeable polypeptide which is then translocated across the membrane.

This method may be applied to study diverse intracellular functions and cellular processes. For instance, it has been used to probe functionally relevant domains of intracellular proteins and to examine protein-protein interactions involved in signal transduction pathways (Lin et al., supra; Lin et al., J. Biol. Chem., 271: 5305-5308, 1996; Rojas et al., J. Biol. Chem., 271: 27456-27461, 1996; Liu et al., Proc. Natl. Acad. Sci. USA, 93: 11819-11824, 1996; Rojas et al., Bioch. Biophys. Res. Commun., 234: 675-680, 1997).

Such techniques may be used in cellular therapy to import proteins producing therapeutic effects. For instance, cells isolated from a patient may be treated with imported therapeutic proteins and then re-introduced into the host organism.

Alternatively, the h region of signal peptides of the present invention could be used in combination with a nuclear localization signal to deliver nucleic acids into cell nucleus. Such oligonucleotides may be antisense oligonucleotides or oligonucleotides designed to form triple helixes, as described in examples 62 and 63 respectively, in order to inhibit processing and/or maturation of a target cellular RNA.

As discussed above, the cDNAs or portions thereof obtained using the 5' ESTs of the present invention can be used for various purposes. The polynucleotides can be used to express recombinant protein for analysis, characterization or therapeutic use; as markers for

tissues in which the corresponding protein is preferentially expressed (either constitutively or at a particular stage of tissue differentiation or development or in disease states); as molecular weight markers on Southern gels; as chromosome markers or tags (when labeled) to identify chromosomes or to map related gene positions; to compare with endogenous DNA sequences in patients to identify potential genetic disorders; as probes to hybridize and thus discover novel, related DNA sequences; as a source of information to derive PCR primers for genetic fingerprinting; for selecting and making oligomers for attachment to a "gene chip" or other support, including for examination for expression patterns; to raise anti-protein antibodies using DNA immunization techniques; and as an antigen to raise anti-DNA antibodies or elicit another immune response. Where the polynucleotide encodes a protein which binds or potentially binds to another protein (such as, for example, in a receptor-ligand interaction), the polynucleotide can also be used in interaction trap assays (such as, for example, that described in Gyuris et al., Cell 75:791-803, 1993, the disclosure of which is hereby incorporated by reference) to identify polynucleotides encoding the other protein with which binding occurs or to identify inhibitors of the binding interaction.

5

10

15

20

25

30

The proteins or polypeptides provided by the present invention can similarly be used in assays to determine biological activity, including in a panel of multiple proteins for high-throughput screening; to raise antibodies or to elicit another immune response; as a reagent (including the labeled reagent) in assays designed to quantitatively determine levels of the protein (or its receptor) in biological fluids; as markers for tissues in which the corresponding protein is preferentially expressed (either constitutively or at a particular stage of tissue differentiation or development or in a disease state); and, of course, to isolate correlative receptors or ligands. Where the protein binds or potentially binds to another protein (such as, for example, in a receptor-ligand interaction), the protein can be used to identify the other protein with which binding occurs or to identify inhibitors of the binding interaction. Proteins involved in these binding interactions can also be used to screen for peptide or small molecule inhibitors or agonists of the binding interaction.

Any or all of these research utilities are capable of being developed into reagent grade or kit format for commercialization as research products.

Methods for performing the uses listed above are well known to those skilled in the art. References disclosing such methods include without limitation *Molecular Cloning*; A

117

Laboratory Manual, 2d ed., Cold Spring Harbor Laboratory Press, Sambrook, Fritsch and Maniatis eds., 1989, and Methods in Enzymology; Guide to Molecular Cloning Techniques, Academic Press, Berger and Kimmel eds., 1987.

Polynucleotides and proteins of the present invention can also be used as nutritional sources or supplements. Such uses include without limitation use as a protein or amino acid supplement, use as a carbon source, use as a nitrogen source and use as a source of carbohydrate. In such cases the protein or polynucleotide of the invention can be added to the feed of a particular organism or can be administered as a separate solid or liquid preparation, such as in the form of powder, pills, solutions, suspensions or capsules. In the case of microorganisms, the protein or polynucleotide of the invention can be added to the medium in or on which the microorganism is cultured.

5

10

15

Although this invention has been described in terms of certain preferred embodiments, other embodiments which will be apparent to those of ordinary skill in the art in view of the disclosure herein are also within the scope of this invention. Accordingly, the scope of the invention is intended to be defined only by reference to the appended claims. All documents cited herein are incorporated herein by reference in their entirety.

	Search characteristic	cteristic	Selection	Selection Characteristics	50
Step	Program	Strand	Parameters	Identity (%)	Length (bp)
miscellanaeous	blastn	both	S=61 X=16	06	17
tRNA	fasta	both		80	9
rRNA	blastn	both	S=108	80	40
mtRNA	blastn	poth	S=108	80	40
Procaryotic	blastn	both	S=144	06	40
Fungal	blastn	both	S=144	06	40
Alu	fasta*	both		70	40
L1	blastn	both	S=72	70	40
Repeats	blastn	both	S=72	20	40
Promoters	blastn	top	S=54 X=16	06	15†
Vertebrate	fasta*	both	S=108	06	30
ESTs	blastn	both	S=108 X=16	06	30
Proteins	blastx¤	top	E = 0.001	•	,

Table 1: Parameters used for each step of EST analysis

use "Quick Fast" Database scanner
 alignement further constrained to begin closer than 10bp to EST\5' end
 using BLOSUM62 substitution matrix

TABLE II

cco m				
SEQ. ID	CATECODY	VON HEIJNE	TISSUE	INTERNAL
<u>NO.</u>	CATEGORY	SCORE	SOURCE	DESIGNATION
ID38	new	15.8	Fetal liver	65-5-1-C9-PU
ID39	new	11.4	Lung (cells)	30-4-2-A11-PU
ID40	new	11.1	Large intestine	83-3-2-H8-PU
ID41	new	10.4	Pancreas	19-8-1-F2
ID42	new	10.4	Liver	22-11-2-H9-PU
ID43	new	9.4	Lung (cells)	30-12-1-H1-PU
ID44	new	9.2	Lung	59-1-3-E7-PU
ID45	new	9.1	Large intestine	83-2-2-D9-PU
ID46	new	9	Lung (cells)	30-2-1-G4-PU
ID47	new	8.9	Colon	23-11-3-C4-PU
ID48	new	8.9	Large intestine	83-2-1-C3-PU
ID49	new	8.6	Lung (cells)	30-3-2-H6-PU
ID50	new	8.6	Lung (cells)	30-13-1-D9-PU
ID51	new	8.6	Colon	23-1-4-E6-PU
ID52	new	8.2	Liver	22-3-3-C4-PU
ID53	new	8	Pancreas	19-4-4-H9
ID54	new	7.7	Lung (cells)	30-8-1-F2-PU
ID55	new	7.5	Lung (cells)	30-6-1-B1-PU
ID56	new	7.5	Lung (cells)	30-6-3-H1-PU
ID57	new	7.5	Colon	23-10-3-F10-PU
ID58	new	7.4	Lung	42-1-1-E3-PU
ID59	new	7.3	Lung	42-3-3-B1-PU
ID60	new	7.3	Lung	42-3-4-B1-PU
ID61	new	7.2	Lung	59-9-2-E6-PU
ID62	new	7	Thyroid	84-4-2-D2-PU
ID63	new	7	Lung (cells)	30-8-3-E3-PU
ID64	new	7	Lung	59-9-4-A10-PU
ID65	new	7	Lung (cells)	30-10-2-A2-PU
ID66	new	6.9	Lung	59-9-1-B9-PU
ID67	new	6.5	Fetal liver	65-4-1-A3-PU
ID68	new	6.5	Lung (cells)	30-2-1-C8-PU
ID69	new	6.4	Colon	23-9-4-F2-PU
ID70	new	6.4	Lung (cells)	30-9-3-A2-PU
ID71	new	6.3	Liver	22-11-2-A9-PU
ID72	new	6.3	Liver	22-13-4-G8-PU
ID73	new	6.2	Liver	22-1-2-A11-PU
ID74	new	6.1	Lung (cells)	30-6-1-D11-PU
ID75	new	5.9	Thyroid	84-4-3-A5-PU
ID76	new	5.8	Lung (cells)	30-5-4-CI-PU
ID77	new	5.7	Liver	22-5-2-A4-PU
ID78	new	5.7	Lung (cells)	30-2-4-B7-PU
ID79	new	5.6	Pancreas	46-3-4-G2-PU
ID80	new	5.6	Thyroid	84-4-3-E9-PU
ID81	new	5.5	Lung (cells)	30-7-2-C7-PU
ID82	new	5.5	Lung (cells)	30-6-3-H11-PU
ID83	new	5.5	Large intestine	83-5-3-C5-PU
ID84	new	5.4	Pancreas	19-1-4-D10
ID85	new	5.3	Lung	59-5-3-A7-PU
ID86	new	5.2	Lung (cells)	30-13-1-G11-PU

WO 99/06439

SEQ. ID		VON HEIJNE	TISSUE	DEFENDIAL
_NO	CATEGORY	SCORE	SOURCE	INTERNAL
		<u> </u>	SOURCE	DESIGNATION
ID87	new	5.2	Lung (cells)	30-7-3-E3-PU
ID88	new	5.2	Lung (cells)	30-9-1-B10-PU
ID89	new	5.1	Lung (cells)	30-8-1-G2-PU
ID90	new	5.1	Lung (cells)	30-13-2-E9-PU
ID91	new	5.1	Thyroid	84-3-3-B4-PU
ID92	new	5	Lung (cells)	30-4-4-D2-PU
ID93	new	4.9	Colon	23-2-4-D1-PU
ID94	new	4.9	Colon	23-9-1-A7-PU
ID95	new	4.9	Lung (cells)	30-11-3-E2-PU
ID96	new	4.9	Fetal liver	65-4-4-C8-PU
ID97	new	4.9	Large intestine	83-2-4-H6-PU
ID98	new	4.8	Lung (cells)	30-5-2-G2-PU
ID99	new	4.8	Liver	22-9-4-B1-PU
ID100	new	4.8	Lung	42-2-2-F2-PU
ID101	new	4.8	Lung (cells)	30-1-1-D5-PU
ID102	new	4.8	Thyroid	84-2-2-G8-PU
ID103	new	4.7	Colon	23-10-4-H5-PU
ID104	new	4.7	Colon	23-8-3-B1-PU
ID105	new	4.7	Lung (cells)	30-8-3-A7-PU
ID106	new	4.6	Lung (cells)	30-11-2-D9-PU
ID107	new	4.5	Lung (cells)	30-6-4-E3-PU
ID108	new	4.5	Large intestine	83-3-2-D3-PU
ID109	new	4.5	Pancreas	19-2-2-E7
ID110	new	4.4	Pancreas	46-1-2-H7-PU
D 111	new	4.4	Colon	23-1-3-C5-PU
ID112	new	4.3	Lung (cells)	30-11-2-E12-PU
ID113	new	4.3	Fetal liver	65-2-3-E3-PU
D114	new	4.3	Colon	23-11-1-G5-PU
ID115	new	4.2	Pancreas	19-2-2-B4
ID116	new	4.2	Lung (cells)	30-4-4-H10-PU
ID117	new	4.2	Lung	42-3-2-F6-PU
ID118	new	4.1	Lung (cells)	30-1-4-G3-PU
ID119	new	4.1	Pancreas	19-4-2-F6
ID120	new	4	Lung	42-3-3-F2-PU
ID121	new	4	Large intestine	83-1-3-H10-PU
ID122	new	4	Lung (cells)	30-11-1-D4-PU
ID123	new	4	Fetal liver	65-5-1-E9-PU
ID124	new	4	Lung (cells)	30-2-3-D4-PU
ID125	new	4	Colon	23-8-4-G8-PU
ID126	new	3.9	Pancreas	19-1-2-D9
ID127	new	3.9	Lung (cells)	30-7-2-D3-PU
ID128	new	3.8	Pancreas	19-3-3-H4
ID129	new	3.7	Fetal liver	65-4-2-F9-PU
D130	new	3.6	Lung (cells)	30-3-3-G 4-PU
ID131	new	3.6	Fetal liver	65-5-2-C3-PU
ID132	new	3.6	Liver	52-3-2- B1-PU
ID133	new	3.6	Large intestine	83-2-2-B12-PU
ID134	new	3.6	Liver	22-10-4-C1-PU
ID135	new	3.6	Thyroid	84-4-1-H8-PU
ID136	new	3.6	Lung (cells)	30-13-4-B11-PU
ID137	new	3.6	Lung (cells)	30-13-1-G12-PU

SEQ. ID		VON HELINE	TISSUE	INTERNAL
_NO	CATEGORY	SCORE	SOURCE	DESIGNATION
		000.00	SOURCE	DESIGNATION
ID138	new	3.5	Pancreas	46-1-4-E11-PU
ID139	new	3.5	Thyroid	84-1-3-C10-PU
ID140	new	3.5	Thyroid	84-4-4-H11-PU
ID141	new	3.5	Lung	59-8-3-A1-PU
ID142	ext-est-not-vrt	7.6	Large intestine	83-4-2-H4-PU
ID143	ext-est-not-vrt	6.6	Lung (cells)	30-2-2-C3-PU
D 144	ext-est-not-vrt	6.6	Thyroid	84-4-1-F7-PU
ID145	ext-est-not-vrt	5.4	Pancreas	19-10-1-C2
ID146	ext-est-not-vrt	5.2	Thyroid	84-5-1-F9-PU
ID147	ext-est-not-vrt	5	Lung	59-9-3-A5-PU
ID148	ext-est-not-vrt	4.7	Lung (cells)	30-7-3-H4-PU
ID149	ext-est-not-vrt	4.5	Lung (cells)	30-11-3-F3-PU
ID150	ext-est-not-vrt	4.4	Lung (cells)	30-12-1-D12-PU
ID151	est-not-ext	16.4	Liver	22-5-3-G5-PU
ID152	est-not-ext	14.4	Large intestine	83-3-2-E8-PU
ID153	est-not-ext	10.3	Liver	52-3-1-B1-PU
ID154	est-not-ext	9.5	Pancreas	19-9-1-C4
ID155	est-not-ext	9.5	Pancreas	19-8-4-F5-PU
ID156	est-not-ext	8.8	Colon	23-2-1-D11-PU
ID157	est-not-ext	8.7	Large intestine	83-4-4-B11-PU
ID158	est-not-ext	8.5	Liver	22-13-3-F7-PU
ID159	est-not-ext	8.1	Lung	42-2-3-A4-PU
ID160	est-not-ext	7.6	Liver	22-10-3-C3-PU
ID161	est-not-ext	7.5	Lung (cells)	30-5-1-B12-PU
ID162	est-not-ext	7.5	Liver	52-1-2-B3-PU
ID163	est-not-ext	6.8	Pancreas	19-8-3-B2
ID164	est-not-ext	6.8	Lung	59-1-3-A4-PU
ID165	est-not-ext	6.8	Thyroid	84-3-1-F10-PU
ID166	est-not-ext	6.1	Lung (cells)	30-12-4-B11-PU
ID167	est-not-ext	5.9	Lung (cells)	30-2-4-B6-PU
ID168	est-not-ext	5.7	Thyroid	84-4-2-D5-PU
ID169	est-not-ext	5.7	Fetal liver	65-2-1-E6-PU
ID170	est-not-ext	5.6	Large intestine	83-5-4-E3-PU
ID171	est-not-ext	5.3	Lung (cells)	30-9-3-E12-PU
ID172	est-not-ext	5	Colon	23-10-3-G8-PU
ID173	est-not-ext	4.9	Lung (cells)	30-10-3-H3-PU
ID174	est-not-ext	4.7	Pancreas	19-3-4-F4
ID175	est-not-ext	4.5	Colon	23-8-3-H9-PU
ID176	est-not-ext	4.3	Large intestine	83-5-4-A4-PU
ID177	est-not-ext	3.8	Pancreas	19-1-3-E11
ID178	est-not-ext	3.8	Lung	59-5-4-A8-PU
ID179	est-not-ext	3.7	Fetal liver	65-4-4-H3-PU
ID180	est-not-ext	3.7	Lung	42-2-1-A1-PU
ID181	est-not-ext	3.6	Liver	22-6-2-C1-PU
ID182	est-not-ext	3.5	Pancreas	46-1-2-B2-PU
ID183	est-not-ext	3.5	Colon	23-12-2-G6-PU
ID184	est-not-ext	3.5	Lung (cells)	30-11-2-H2-PU
			wang (cens)	20-11-2-17-10

122

TABLE III

SEQ. ID	
_NÔ.	SIGNAL PEPTIDE
ID38	MMWRPSVLLLLLLRHGAQG
ID39	MGKICKNWVSFLDNVLLLILFLYGLCSG
ID40	MLTVALLALLCASASGNA
ID41	MVLLLCLSCLIFS
ID42	MPVPALCLLWALAMVTRPASA
ID43	MHLRGSHTYPSCPSSELRLDSLWQHHRQLLPLWVFLPLSLG
ID44	MPVPASWPHLPSPFLLMTLLLGGLTG
ID45	MAQRCVCVLALVAMLLLVFPTVS
ID46	MDYLISFLLLLLLP
ID47	MATTVPDGCRNGLKSKYYRLCDKAEAWGIVLETVATAGVVTSVAFMLTLPILVCKV
	QDSNRRKMLPTQFLFLLGVLGIFGLTFA
ID48	MESGLSWLFLVIFIKGVQC
ID49	MSGTSVLLHVAFLPGRFG
ID50	MLQGLLPVSLLLSVAVS
ID51	MHICHVSLLLQLCSS
ID52	MIFADRTHSSAFTLMRSYSLLLCSLLFSFPFLC
ID53	MAFLPSWVCVLVGSFSASLA
ID54	MFLVSCVICTGSFA
ID55	MKKTGDGGTLSTERIGGAALLSLLLKRMKMTLMIPLLLLTPITA
ID56	MGFFLPHGISDAXILLAGWCPDTRA
ID57	MWLRPGSCWSTREPRRAPRTSASSLSSFLGPSAVCTLLSSHPASRC
ID58	MSEGMVTLLTFSCLWTDDSFMSXLNVLFLLSLFCRLYHG
ID59	MLILGLPLCRPLWI
ID60	MYTYFFVLCXLSHFILLVLPCLIFS
ID61	MDSRVSSPEKQDKENFVGVNNKRLGVCGWILFSLSFLLVIIT
ID62	MCILFCVVLCLSPTSY
ID63	MHRGDIETLLCLGSSCCQC
ID64	MFLKSGAGLSSCLLPLCWL
ID65	MANAIIKKPCAMPAQPHTGNLLWPPLVMVWLGLLPLFS
ID66	MSPPPLLQPLLLLPLLNV
ID67	
ID68	MIPIYQNKSQTDSHCSLSHKGLAFLKVWLILIGLFSLTGLVA MALPGIHLLSGSTCPGPCSC
ID69	
ID70	MPSETLWEIAKAEVEKRGINGXXGDGAEIALIPLFSTXAFA
ID71	MEWLRPSQISFYPGYSKERLRLVLLCMSLTFLALSTL
	MKAIHLTLLALLSVNTG
ID72	MDVSASKPVAESWSPGSLPLALTLSLSTS
ID73	MGVRVGVSLRAWCVFIQTALLGLPXAWA
ID74	MISIPRSFFLLLCIPFLTLL
ID75	MTMQRSRSSSWTSCNSWTLVLMSPEWALL
ID76	MITLPQTSSLLCSLMASISPTLT
ID77	MLRTCYVLCSQAGPPSRGWQSLSFDGGAFHLKGTGELTRALLVLRLCAWPPLVTHGL
man.	LLQAWS
ID78	MICSPFSGFAPCQALGTLGVGCHFFHLALG
ID79	MCNPEEAALXGLEEVFSATLAHVNSLVLQPLLPAAPDPSDPWGRECLRLLQQLHKSSQQL
TD00	WEVTEESLHSLQERLRYPDSTGLESLLLLRGADRVLQA
ID80	MDKLIPSLSSQENRKASHTLHKARNKQHCGGFLLVIHWVMCPSLS
ID81	MSXLLPVVLASPPVGHG
ID82	MVLLTMIARVADG
ID83	MFHIAFSEALPVDIFKTQPNCHEAFSMKAIHITRIRSGLCLLELLFVPLLCFL

SEQ. ID	
NO	SIGNAL PEPTIDE
_	
ID84	MMHCTPSGSAAVSLLTETVLPLAFP
ID85	MTRPFWASCSTWATSRISCAFSLASSTA
ID86	MVTHLIRGVVLQGSCCLIQWPELSFS
ID87	MYMWSKLLVAFRVFLGLFS
ID88	MSSRNCFFPSFLFGLYSFRAVDS
ID89	MYMNTCLYLHVYVLTCSG
ID90	MSCRQPTPTQCSLLPNDNRVSTRGGDSAGRHRQVPQVALSASLPQCSLG
ID91	MITGCTKPTAGVVVLQGSRA
ID92	MGLDLILSFSSSSP
ID93	MREDNEHERNVPSGVENVKEEGGDEDLSWGDEGCQVLRHRLRVCRKVGLLDRLCA
	LTSLCSP
ID94	MGKRAGAVVSSWAXCSLG
ID95	MQSTSNHLWLLSDILGQGATA
ID96	MKKLRPSQEQLNCPEPQLADGRAGIRLLVTWLQPAPLLCLSGLELEPSA
ID97	MWSHLNRLLFWSIFSSVTC
ID98	MLALRDLGMGKREGEELIQAEARCLVETFQGTEGRPFDPSLLLAQATSNVVC
ID99	MLSVGASTSLCGCLRQLRC
ID100	MFQQMYVLLSQFLYPLAYP
ID101	MTSHFCXIGFLSYTTS
ID102	MICSLTPFRSLTNVLLSGSLLRSLC
ID103	MEPPGRSSSLPFSPPALTLTFLPPSPT
ID104	MDKLKKVLSGQDTEDRSGLSEVVEASSLSWSTRIKGFIACFAIGILCSLLGTVLL
ID105	MYSRHTVKLKQGLGMVCIFSLRLQA
ID106	MYPSLLVDYFPSLLLYSLPLNIIG
ID107	MATTVPDGCRNGLKSKYYRLCDKAEAWGIVLETVATAGVVTSVAFMXTLPILVCKVQ
	DSNRRKMLPTQFLFLLGVLG
ID108	MRLQHLDHLFFSGVVLG
ID109	MPLPKPSFSNNHLIRLITVAFGLYNPSLCHA
ID110	MEPITFTARKHLLPNEVSVDFGLQLVGSLPVHSLTTMPMLPWVVAEVRRLSRQSTRKEPV
	TXQXRLCVSPSGLRC
ID111	MGCLWGLALPLFFFCWEVGVSGSSA
ID112	MKQNTDPYLCHISLLDVTQQ
ID113	MVTYFNFTFKPFCILASIIVPTLISLLSSPNTPSA
ID114	MESGGRPSLCQXILLGTTSVVTA
ID115	MEAQQAQKSAEQPEQKAATEVSXELSESQVHMMAAAVADGTRA
ID116	MPLNSVIWFGSVXPCIS
ID117	MLQQLDSISLRRRETANFLDFANLADLTLA
ID118	
ID119	MCYLAELSLTTFXXGYIVTSRATTTTTLAIQPGLPFTTLSNLSLPSQT MSISLSSLILLPIWINM
ID120	
1120	MDRDLLRQSLNCHGSSLLSLLRSEQQDNPHFRSLLGSAAEPARGPPPQHPLQGRKEKRVD
ID121	NIEIQKFISKKADLLFALSWKSDAPA
ID121	MVLATLVTXXNASCSFA
ID122	MMIWKRLIILKVLLNQTCQT
ID123	MDAGKAGQTLKTHCSAQRPDVCRWLSPFILSCCVYFCLWIPEDQLSWFAALVKCLPV
FD124	LCLA
ID124	MQQRGAAGSRGCALFPLLGVLFFQGVYI
ID125	MLGTHIYVSLWIILFSSPHLIYWYVLLILSFP
ID126	MSIYNLFLNLHGFLGHLLS
ID127	MCMQVDLAFSFPPACVCMCTXSCYS
ID128	MAPGEKESGEGPAKSALRKIRTATLVISLARG
ID129	MEPKRGRMWXFEIEDSCIYODIPSFVI I YPI I HI FYOHI CFP

SEQ. ID	
_NO	SIGNAL PEPTIDE
	<u> </u>
ID130	MEFCSVLQRCLFSFVTS
ID131	MAESQIYVLLFFLLMKFS
ID132	MQTNNACSLSSGPLQINA
ID133	MGQNNASFHCPCLKVLMGLLCNQTAA
ID134	MLPLLSVMWSPIAP
ID135	MWLNCGGLQRWITCPPTFHGCRA
ID136	MWQGCNCSQLSETAVDQEQLGVLTFILQRTTC
ID137	MCLPHPQVVSSNFHILIFLLPTKMLVTLLASKSPSCP
ID138	MHLAVLFXFSDCCRKXLSSGQLYSIVSSLSNEHVLSAGFDINTPDNLGRTCLHAAASGGN
10100	VECLNLLLSSGADL
ID139	MSFQWCGWQWGLHDCFLSVFQVLS
ID140	MKVHMHTKFCLICLLTFIFH
ID141	MSFNLQSSKKLFIFLGKSLFSLLEA
ID142	MDLMCRKVKHLLFFLLLVAAPRWVVS
ID143	
ID143	MELKSPEEEVVAALPEGMRPDSNLYGFPWELVICAAVVGFFAVLFFLWRSFX
ID145	MELSDVTLIEGVGNEVMVVAGVVVLILALVLAWLSTYVA MIARRNPEPLRFLPDEARSLPPPKLTDPRLLYIGFLGYCSG
D146	MPPGPWESCFWVGGLILWLSVGSS
ID147	MCARALLIACSSRG
ID148	
ID149	MGDERPHYYGKHGTPQKYDPTFKGPIYNRGCTDIICCVFLLLAIVG MAQRLLLRRFLASVIS
ID150	
ID151	MESGGRPSLCQFILLGTTSVVTA MALSSONIA ACLULIUM ASSTER
ID152	MALSSQIWAACLLLLLLASLTSG MGVPPPOPWALCLLLFLLASLTSG
	MGVPRPQPWALGLLFLLPGSLG
ID153	MKVVPSLLLSVLLAQVWL
ID154	MLSITVLAALLACASS
ID155 ID156	MLGITVLAALLACASS
ID157	MAGNGESEPDRLHLLTGHRVKGEFQLLLPLLSLPVTTP
	MLWWLVLLLLPTLK MAPOSI PSSPNAPI CHARLES AND
ID158	MAPQSLPSSRMAPLGMLLGLLMAACFTFC
D159	MMLHSALGLCLLLVTVSSNLAIA MCTCVCARCVCLSUTT CLACULARIA
ID160	MCTGKCARCVGLSLITLCLVCIVANA
ID161	MDILVPLLQLLVLLLTLPLHLMA
ID162	MPFLVLFSFFNIALC
ID163	MQQRGLAIVALAVCAALHA
D164	MRKTRLWGLLWMLFVSELRA
ID165	MVGAMWKVIVSLVLLMPGPCDG
ID166	MIHLRIIQRCYMAGLENKKNVVFEAKQICIGILVLPFIRC
ID167	MAGSPTCLTLIYILWQLTGSAA
ID168	MGKKGKVGKSRRDKFYHLAKETGYRSRSAFKLIQLNRRFQFLQKARALLDLCAAPXGWL
ID169	MPLSDFILALKDNPYFGAGFGLVXVGTALALA
ID170	MEFGLSWVFLVAIIKGVQC
ID171	MILRKRSCSLFSSLPIFLTWA
ID172	MKNGLMFVKLVNPCSG
ID173	MEAVVFVFSLLDCCA
ID174	MTGFLLPPASRGTRRSCSRSRKRQTRRRRNPSSFVASCPTLLPFACVPGASXTTLA
ID175	MCGNTMSVPLLTDAATVSG
ID176	MXXXXERRTSPHVMADQSSTRNEDFLKKTWSLWRLQWLKDASC
ID177	MFLLLNCIVAVSQN
ID178	MLLVSAAPLGFGQG
ID179	MLRIALTLIPSMLSRA

SEQ. ID NO.	SIGNAL PEPTIDE
ID180	MTLGGRLPGLRCSVPGVAA
ID181	MAFTLXSLLQAALL
ID182	MRPLAGGLLKVVFMVFASLXA
ID183	MFEEPEWAEAAPVAAGLGPVIS
ID184	MHIYTGIKYIAI YNIIYVCGIOG

Minimum signal peptide score	false positive rate	false negative rate	proba(0.1)	proba(0.2)
3.5	0.121	0.036	0.467	0.664
4	0.096	0.06	0.519	0.708
4.5	0.078	0.079	0.565	0.745
5	0.062	0.098	0.615	0.782
5.5	0.05	0.127	0.659	0.813
6	0.04	0.163	0.694	0.836
6.5	0.033	0.202	0.725	0.855
7	0.025	0.248	0.763	0.878
7.5	0.021	0.304	0.78	0.889
8	0.015	0.368	0.816	0.909
8.5	0.012	0.418	0.836	0.92
9	0.009	0.512	0.856	0.93
9.5	0.007	0.581	0.863	0.934
10	0.006	0.679	0.835	0.919

TABLE IV

125/2

Minimum signal peptide score	All ESTs	New ESTs	ESTs matching public EST closer than 40 bp from beginning	ESTs extending known mRNA more than 40 bp	ESTs extending public EST more than 40 bp
3.5	2674	947	599	23	150
4	2278	784	499	23	126
4.5	1943	647	425	22	112
5	1657	523	353	21	96
5.5	1417	419	307	19	80
6	1190	340	238	18	68
6.5	1035	280	186	18	60
7	893	219	161	15	48
7.5	753	173	132	12	36
8	636	133	101	11	29
8.5	543	104	83	8	26
9	456	81	63	6	24
9.5	364	57	48	6	18
10	303	47	35	6	15

TABLE V

Tissue	All ESTs	New ESTs	ESTs matching public EST closer than 40 bp from beginning	ESTs extending known mRNA more than 40 bp	ESTs extending public EST more than 40 bp
Brain	329	131	75	3	24
Cancerous prostate	134	40	37	1	6
Cerebellum	17	9	1	0	6
Colon	21	11	4	0	ol
Dystrophic muscle	41	18	8	Ō	1
Fetal brain	70	37	16	0	1
Fetal kidney	227	116	46	1	19
Fetal liver	13	7	2	0	0
Heart	30	15	7	0	1
Hypertrophic prostate	86	23	22	2	2
Kidney	10	7	3	0	ō
Large intestine	21	8	4	0	1
Liver	. 23	9	6	0	o
Lung	24	12	4	0	1
Lung (cells)	57	38	6	0	4
Lymph ganglia	163	60	23	2	12
Lymphocytes	23	6	4	0	2
Muscle	33	16	6	0	4
Normal prostate	181	61	45	7	11
Ovary	90	57	12	1	2
Pancreas	48	11	6	0	1
Placenta	24	5	1	0	o
Prostate	34	16	4	0	2
Spieen	56	28	10	0	1
Substantia nigra	108	47	27	1	6
Surrenals	15	3	3	1	o
Testis	131	68	25	1	8
Thyroid	17	8	2	0	2
Umbilical cord	55	17	12	1	3
Uterus	28	15	3	0	2
Non tissue-specific	568	48	177	2	28
Total	2677	947	601	23	150

TABLE VI

125/4

Description of Transcription Factor Binding Sites present on promoters isolated from SignalTag sequences Promoter sequence P13H2 (646 bp):

Matrix	Position	Orientation	Score	Length	Sequence
CMYB_01	-502	•	0.983	9	TGTCAGTTG
MYOD_Q6	-501	•	0.961	10	CCCAACTGAC
S8_01	-444	•	0.960	11	AATAGAATTAG
S8_01	-425	•	0.966	11	AACTAAATTAG
DELTAEF1_01	-390	•	0.960	11	GCACACCTCAG
GATA_C	-384	•	0.964	11	AGATAAATCCA
CMYB_01	-349	+	0.958	9	CTTCAGTTG
GATA1_02	-343	+	0.959	14	TTGTAGATAGGACA
GATA_C	-339	+	0.953	11	AGATAGGACAT
TAL1ALPHAE47_01	-235	•	0.973	16	CATAACAGATGGTAAG
TAL1BETAE47 01	-235	+	0.983	16	CATAACAGATGGTAAG
TAL1BETAITF2_01	-235	+	0.978	16	CATAACAGATGGTAAG
MYOD_Q8	-232	•	0.954	10	ACCATCTGTT
GATA1_04	-217	-	0.953	13	
IK1_01	-126	•	0.963	13	TCAAGATAAAGTA
IK2_01	-126	•	0.985		AGTTGGGAATTCC
CREL_01	-123	•	0.962	12	AGTTGGGAATTC
GATA1_02	-96	•	0.952	10	TGGGAATTCC
SRY_02	-41	•		14	TCAGTGATATGGCA
E2F_02	-33	•	0.951	12	TAAAACAAAACA
MZF1_01	-55 -5	•	0.957	8	TTTAGCGC
11401 1_01	~	•	0.975	8	TGAGGGGA

Promoter sequence P16B4 (861bp):

Matrix	Position	Orientation	Score	Length	Sequence
NFY_Q6	-748	-	0.956	11	GGACCAATCAT
MZF1_01	-738	+	0.962	8	CCTGGGGA
CMYB_01	-684	+	0.994	9	TGACCGTTG
VMYB_02	-682	-	0.985	9	TCCAACGGT
STAT_01	-673	+	0.968	9	TTCCTGGAA
STAT_01	-673	•	0.951	9	TTCCAGGAA
MZF1_01	-556	-	0.956	8	TTGGGGGA
IK2_01	-451	+	0.965	12	GAATGGGATTTC
MZF1_01	-424	+	0.986	8	AGAGGGGA
SRY_02	-398	•	0.955	12	GAAAACAAAACA
MZF1_01	-216	+	0.960	8	GAAGGGGA
MYOD_Q6	-190	+	0.981	10	AGCATCTGCC
DELTAEF1_01	-176	+	0.958	11	TCCCACCTTCC
S8_01	5	-	0.992	11	GAGGCAATTAT
MZF1_01	16	-	0.986	8	AGAGGGGA

Promoter sequence P29B6 (655 bp):

Matrix	Position	Orientation	Score	Length	Sequence
ARNT_01	-311	+	0.964	16	GGACTCACGTGCTGCT
NMYC_01	-309	+	0.965	12	ACTCACGTGCTG
USF_01	-309	+	0.985	12	ACTCACGTGCTG
USF_01	-309	•	0.985	12	CAGCACGTGAGT
NMYC_01	-309	-	0.956	12	CAGCACGTGAGT
MYCMAX_02	-309	•	0.972	12	CAGCACGTGAGT
USF_C	-307	•	0.997	8	TCACGTGC
USF_C	-307	•	0.991	8	GCACGTGA
MZF1_01	-292	-	0.968	8 .	CATGGGGA
ELK1_02	-105	+	0.963	14	CTCTCCGGAAGCCT
CETS1P54_01	-102	+	0.974	10	TCCGGAAGCC
AP1_Q4	-42	-	0.963	11	AGTGACTGAAC
AP1FJ_Q2	-42	•	0.961	11	AGTGACTGAAC
PADS_C	45	+	1.000	9	TGTGGTCTC

TABLE VII

15

CLAIMS

- A purified or isolated nucleic acid comprising the sequence of one of SEQ ID
 NOs: 38-184 or comprising a sequence complementary thereto.
 - 2. The nucleic acid of Claim 1, wherein said nucleic acid is recombinant.
- 3. A purified or isolated nucleic acid comprising at least 10 consecutive bases of the sequence of one of SEQ ID NOs: 38-184 or one of the sequences complementary thereto.
- 4. A purified or isolated nucleic acid comprising at least 15 consecutive bases of one of the sequences of SEQ ID NOs: 38-184 or one of the sequences complementary thereto.
 - 5. The nucleic acid of Claim 4, wherein said nucleic acid is recombinant.
 - 6. A purified or isolated nucleic acid of at least 15 bases capable of hybridizing under stringent conditions to the sequence of one of SEQ ID NOs: 38-184 or one of the sequences complementary to the sequences of SEQ ID NOs: 38-184.
 - 7. The nucleic acid of Claim 6, wherein said nucleic acid is recombinant.
 - 8. A purified or isolated nucleic acid encoding a human gene product, said human gene product having a sequence partially encoded by one of the sequences of SEQ ID NO: 38-184.
- A purified or isolated nucleic acid having the sequence of one of SEQ ID
 NOs: 38-184 or having a sequence complementary thereto.
 - 10. A purified or isolated nucleic acid comprising the nucleotides of one of SEQ ID NOs: 38-184 which encode a signal peptide.
- 11. A purified or isolated polypeptides comprising a signal peptide encoded by one of the sequences of SEQ ID NOs: 38-184.
 - 12. A vector encoding a fusion protein comprising a polypeptide and a signal peptide, said vector comprising a first nucleic acid encoding a signal peptide encoded by one of the sequences of SEQ ID NOs: 38-184 operably linked to a second nucleic acid encoding a polypeptide.
- 30 13. A method of directing the extracellular secretion of a polypeptide or the insertion of a polypetide into the membrane comprising the steps of:

10

20

WO 99/06439 PCT/IB98/01233 127

obtaining a vector according to Claim 12; and

introducing said vector into a host cell such that said fusion protein is secreted into the extracellular environment of said host cell or inserted into the membrane of said host cell.

- A method of importing a polypeptide into a cell comprising contacting said 14. cell with a fusion protein comprising a signal peptide encoded by one of the sequences of SEQ ID NOs: 38-184 operably linked to said polypeptide.
- A method of making a cDNA encoding a human secretory protein that is 15. partially encoded by one of SEQ ID NOs 38-184, comprising the steps of:

obtaining a cDNA comprising one of the sequences of SEQ ID NOs: 38-184;

contacting said cDNA with a detectable probe comprising at least 15 consecutive nucleotides of said sequence of SEQ ID NO: 38-184 or a sequence complementary thereto under conditions which permit said probe to hybridize to said cDNA;

identifying a cDNA which hybridizes to said detectable probe; and isolating said cDNA which hybridizes to said probe.

- 15 16. An isolated or purified cDNA encoding a human secretory protein, said human secretory protein comprising the protein encoded by one of SEQ ID NOs 38-184 or a fragment thereof of at least 10 amino acids, said cDNA being obtainable by the method of Claim 15.
 - The cDNA of Claim 16 wherein said cDNA comprises the full protein coding 17. sequence partially included in one of the sequences of SEQ ID NOs: 38-184.
 - 18. A method of making a cDNA comprising one of the sequences of SEQ ID NOs: 38-184, comprising the steps of:

contacting a collection of mRNA molecules from human cells with a first primer capable of hybridizing to the polyA tail of said mRNA:

25 hybridizing said first primer to said polyA tail:

reverse transcribing said mRNA to make a first cDNA strand;

making a second cDNA strand complementary to said first cDNA strand using at least one primer comprising at least 15 nucleotides of one of the sequences of SEQ ID NOs 38-184; and

30 isolating the resulting cDNA comprising said first cDNA strand and said second cDNA strand.

10

15

30

- 19. An isolated or purified cDNA encoding a human secretory protein, said human secretory protein comprising the protein encoded by one of SEQ ID NOs 38-184 or a fragment thereof of at least 10 amino acids, said cDNA being obtainable by the method of Claim 18.
- 20. The cDNA of Claim 19 wherein said cDNA comprises the full protein coding sequence partially included in one of the sequences of SEQ ID NOs: 38-184.
- 21. The method of Claim 18, wherein the second cDNA strand is made by: contacting said first cDNA strand with a first pair of primers, said first pair of primers comprising a second primer comprising at least 15 consecutive nucleotides of one of the sequences of SEQ ID NOs 38-184 and a third primer having a sequence therein which is included within the sequence of said first primer;

performing a first polymerase chain reaction with said first pair of nested primers to generate a first PCR product;

contacting said first PCR product with a second pair of primers, said second pair of primers comprising a fourth primer, said fourth primer comprising at least 15 consecutive nucleotides of said sequence of one of SEQ ID NO:s 38-184, and a fifth primer, said fourth and fifth primers being capable of hybridizing to sequences within said first PCR product; and

performing a second polymerase chain reaction, thereby generating a second PCR product.

- 22. An isolated or purified cDNA encoding a human secretory protein, said human secretory protein comprising the protein encoded by one of SEQ ID NOs 38-184, or a fragment thereof of at least 10 amino acids, said cDNA being obtainable by the method of Claim 21.
- 23. The cDNA of Claim 22 wherein said cDNA comprises the full protein coding sequence partially included in one of the sequences of SEQ ID NOs: 38-184.
 - 24. The method of Claim 18 wherein the second cDNA strand is made by: contacting said first cDNA strand with a second primer comprising at least 15 consecutive nucleotides of the sequences of SEQ ID NOs: 38-184;

hybridizing said second primer to said first strand cDNA; and extending said hybridized second primer to generate said second cDNA strand.

- 25. An isolated or purified cDNA encoding a human secretory protein, said human secretory protein comprising the protein partially encoded by one of SEQ ID NOs 38-184 or comprising a fragment thereof of at least 10 amino acids, said cDNA being obtainable by the method of Claim 24.
- 5 26. The cDNA of Claim 25, wherein said cDNA comprises the full protein coding sequence partially included in of one of the sequences of SEQ ID NOs: 38-184.
 - 27. A method of making a protein comprising one of the sequences of SEQ ID NO: 185-331, comprising the steps of:

obtaining a cDNA encoding the full protein sequence partially included in one of the sequences of sequence of SEQ ID NO: 38-184;

inserting said cDNA in an expression vector such that said cDNA is operably linked to a promoter;

introducing said expression vector into a host cell whereby said host cell produces the protein encoded by said cDNA; and

isolating said protein.

25

- 28. An isolated protein obtainable by the method of Claim 27.
- 29. A method of obtaining a promoter DNA comprising the steps of:

obtaining DNAs located upstream of the nucleic acids of SEQ ID NO: 38-184 or the sequences complementary thereto;

screening said upstream DNAs to identify a promoter capable of directing transcription initiation; and

isolating said DNA comprising said identified promoter.

- 30. The method of Claim 29, wherein said obtaining step comprises chromosome walking from said nucleic acids of SEQ ID NO: 38-184 or sequences complementary thereto.
- 31. The method of Claim 30, wherein said screening step comprises inserting said upstream sequences into a promoter reporter vector.
 - 32. The method of Claim 30, wherein said screening step comprises identifying motifs in said upstream DNAs which are transcription factor binding sites or transcription start sites.
- 30 An isolated promoter obtainable by the method of Claim 32.

- 34. An isolated or purified protein comprising one of the sequences of SEQ ID NO: 185-331.
- 35. In an array of discrete ESTs or fragments thereof of at least 15 nucleotides in length, the improvement comprising inclusion in said array of at least one of the sequences of SEQ ID NOs: 38-184, or one of the sequences complementary to the sequences of SEQ ID NOs: 38-184, or a fragment thereof of at least 15 consecutive nucleotides.
- 36. The array of Claim 35 including therein at least two of the sequences of SEQ ID NOs: 38-184, the sequences complementary to the sequences of SEQ ID NOs: 38-184, or fragments thereof of at least 15 consecutive nucleotides.
- 10 37. The array of Claim 35 including therein at least five of the sequences of SEQ ID NOs: 38-184, the sequences complementary to the sequences of SEQ ID NOs: 38-184, or fragments thereof of at least 15 consecutive nucleotides.

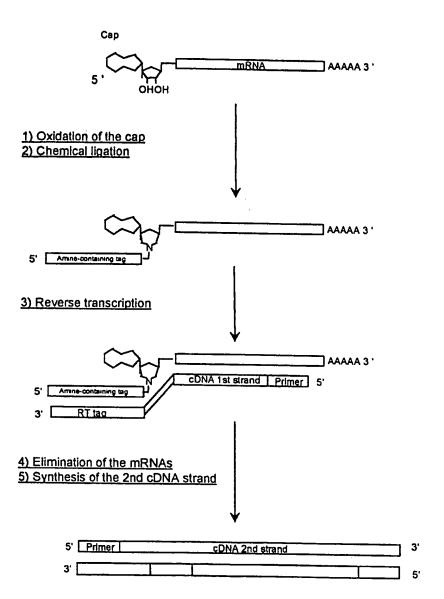


Figure 1

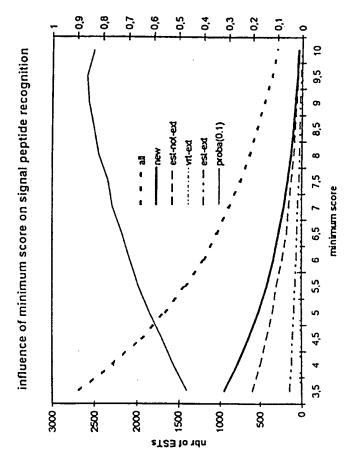


Figure 2

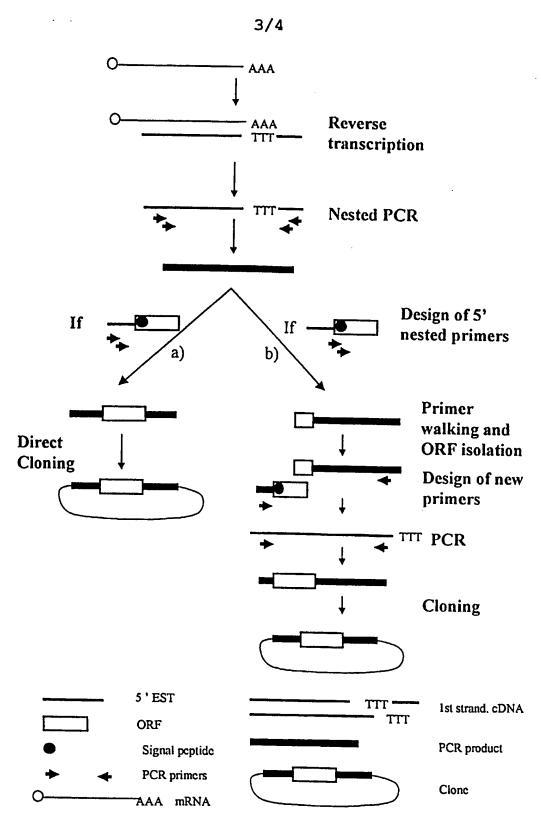
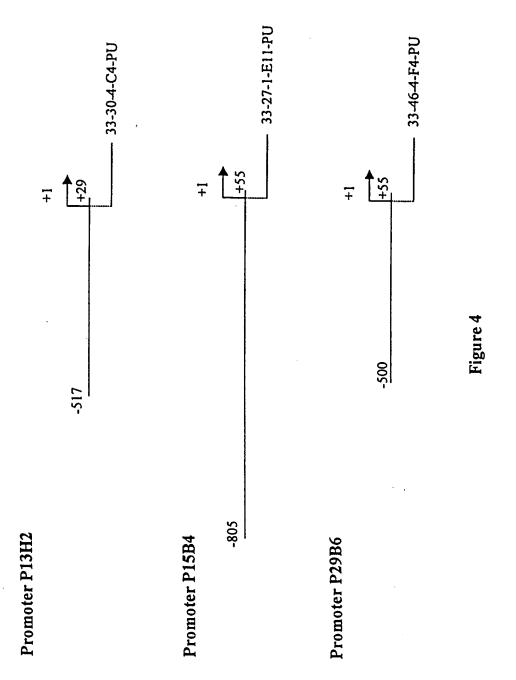


Figure 3



SEQUENCE LISTING

- (1) GENERAL INFORMATION:
 - (i) APPLECANT:
 - (A) NAME : GENSET SA
 - (B) STREET :24, RUE ROYALE
 - (C) CITY: PARIS
 - (E) COUNTRY : FRANCE
 - (F) POSTAL CODE (ZIP): 75008
 - (ii) TITLE OF INVENTION: 5' ESTs FOR SECRETED PROTEINS EXPRESSED IN ENDODERM
 - (iii) NUMBER OF SEQUENCES: 331
 - (v) COMPUTER READABLE FORM:
 - (A) MEDIUM TYPE: Floppy Disk
 - (B) COMPUTER: IBM PC compatible
 - (C) OPERATING SYSTEM: Win95
 - (D) SOFTWARE: Word
- (2) INFORMATION FOR SEQ ID NO: 1:
 - (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 47 base pairs
 - (B) TYPE: NUCLEIC ACID
 - (C) STRANDEDNESS: SINGLE
 - (D) TOPOLOGY: LINEAR
 - (ii) MOLECULE TYPE: Other nucleic acid
 - (ix) FEATURE:
 - (A) NAME/KEY: Cap
 - (B) LOCATION: 1
 - (D) OTHER INFORMATION: m7Gppp added to 1
 - (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 1:

GGCAUCCUAC UCCCAUCCAA UUCCACCCUA ACUCCUCCCA UCUCCAC

- (2) INFORMATION FOR SEQ ID NO: 2:
 - (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 46 base pairs
 - (B) TYPE: NUCLEIC ACID
 - (C) STRANDEDNESS: SINGLE
 - (D) TOPOLOGY: LINEAR
 - (ii) MOLECULE TYPE: Other nucleic acid
 - (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 2:

(2)	INFORMATION FOR SEQ ID NO: 3:	
	(i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 25 base pairs (B) TYPE: NUCLEIC ACID (C) STRANDEDNESS: SINGLE (D) TOPOLOGY: LINEAR	
	(ii) MOLECULE TYPE: Other nucleic acid	
	(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 3:	
ATC	CAAGAATT CGCACGAGAC CATTA	25
(2)	INFORMATION FOR SEQ ID NO: 4:	
	(i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 25 base pairs (B) TYPE: NUCLEIC ACID (C) STRANDEDNESS: SINGLE (D) TOPOLOGY: LINEAR	
	(ii) MOLECULE TYPE: Other nucleic acid	
	(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 4:	
TAA	TGGTCTC GTGCGAATTC TTGAT	25
(2)	INFORMATION FOR SEQ ID NO: 5:	
	(i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 25 base pairs (B) TYPE: NUCLEIC ACID (C) STRANDEDNESS: SINGLE (D) TOPOLOGY: LINEAR	
	(ii) MOLECULE TYPE: Other nucleic acid	
	(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 5:	
CCG	ACAAGAC CAACGTCAAG GCCGC	25
(2)	INFORMATION FOR SEQ ID NO: 6:	
	(i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 25 base pairs (B) TYPE: NUCLEIC ACID (C) STRANDEDNESS: SINGLE (D) TOPOLOGY: LINEAR	

WO 99/06439	3	PCT/IB98/01233
(ii) MOLECULE TYPE: Other nucle	ic acid	
(xi) SEQUENCE DESCRIPTION: SEQ	ID NO: 6:	
TCACCAGCAG GCAGTGGCTT AGGAG		25
(2) INFORMATION FOR SEQ ID NO: 7:		
(i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 25 base pairs (B) TYPE: NUCLEIC ACID (C) STRANDEDNESS: SINGLE (D) TOPOLOGY: LINEAR		
(ii) MOLECULE TYPE: Other nuclei	ic acid	
(xi) SEQUENCE DESCRIPTION: SEQ	ID NO: 7:	
AGTGATTCCT GCTACTTTGG ATGGC		25
(2) INFORMATION FOR SEQ ID NO: 8:		
(i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 25 base pairs (B) TYPE: NUCLEIC ACID (C) STRANDEDNESS: SINGLE (D) TOPOLOGY: LINEAR		
(ii) MOLECULE TYPE: Other nuclei	ic acid	
(xi) SEQUENCE DESCRIPTION: SEQ I	ID NO: 8:	
GCTTGGTCTT GTTCTGGAGT TTAGA		25
(2) INFORMATION FOR SEQ ID NO: 9:		
(i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 25 base pairs (B) TYPE: NUCLEIC ACID (C) STRANDEDNESS: SINGLE (D) TOPOLOGY: LINEAR		
(ii) MOLECULE TYPE: Other nuclei	c acid	
(xi) SEQUENCE DESCRIPTION: SEQ I	ID NO: 9:	
TCCAGAATGG GAGACAAGCC AATTT		25

(2) INFORMATION FOR SEQ ID NO: 10:

	(i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 25 base pairs (B) TYPE: NUCLEIC ACID (C) STRANDEDNESS: SINGLE (D) TOPOLOGY: LINEAR	
	(ii) MOLECULE TYPE: Other nucleic acid	
	(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 10:	
AGGG	AGGAGG AAACAGCGTG AGTCC	25
(2)	INFORMATION FOR SEQ ID NO: 11:	
	(i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 25 base pairs (B) TYPE: NUCLEIC ACID (C) STRANDEDNESS: SINGLE (D) TOPOLOGY: LINEAR	
	(ii) MOLECULE TYPE: Other nucleic acid	
	(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 11:	
ATGG	GAAAGG AAAAGACTCA TATCA	25
(2)	INFORMATION FOR SEQ ID NO: 12:	
	(i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 25 base pairs (B) TYPE: NUCLEIC ACID (C) STRANDEDNESS: SINGLE (D) TOPOLOGY: LINEAR	
	(ii) MOLECULE TYPE: Other nucleic acid	
	(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 12:	
AGCA	GCAACA ATCAGGACAG CACAG	25
(2)	INFORMATION FOR SEQ ID NO: 13:	
	(i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 25 base pairs (B) TYPE: NUCLEIC ACID (C) STRANDEDNESS: SINGLE (D) TOPOLOGY: LINEAR	
	(ii) MOLECULE TYPE: Other nucleic acid	
	(xi) SEQUENCE DESCRIPTION: SEC ID NO: 13:	

VO 99/06439	•		PCT/IB98/01233
·		5	

ATCAAGAATT CGCACGAGAC CATTA

(2) INFORMATION FOR SEQ ID NO: 14:

(i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 67 base pairs (B) TYPE: NUCLEIC ACID (C) STRANDEDNESS: SINGLE (D) TOPOLOGY: LINEAR	
(ii) MOLECULE TYPE: Other nucleic acid	
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 14:	
ATCGTTGAGA CTCGTACCAG CAGAGTCACG AGAGAGACTA CACGGTACTG GTTTTTTTT	60
TTTTVN	67
(2) INFORMATION FOR SEQ ID NO: 15:	
(i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 29 base pairs (B) TYPE: NUCLEIC ACID (C) STRANDEDNESS: SINGLE (D) TOPOLOGY: LINEAR	
(ii) MOLECULE TYPE: Other nucleic acid	
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 15:	
CCAGCAGAGT CACGAGAGAG ACTACACGG	29
(2) INFORMATION FOR SEQ ID NO: 16:	
 (i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 25 base pairs (B) TYPE: NUCLEIC ACID (C) STRANDEDNESS: SINGLE (D) TOPOLOGY: LINEAR 	
(ii) MOLECULE TYPE: Other nucleic acid	
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 16:	
CACGAGAGAG ACTACACGGT ACTGG	25
(2) INFORMATION FOR SEQ ID NO: 17:	

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 526 base pairs
- (B) TYPE: NUCLEIC ACID
- (C) STRANDEDNESS: DOUBLE
- (D) TOPOLOGY: LINEAR
- (ii) MOLECULE TYPE: CDNA
- (vi) ORIGINAL SOURCE:
 - (A) ORGANISM: Homo Sapiens
 - (F) TISSUE TYPE: Lymph ganglia
- (ix) FEATURE:
 - (A) NAME/KEY: other
 - (B) LOCATION: complement (261..376)
 - (C) IDENTIFICATION METHOD: blastn
 - (D) OTHER INFORMATION: identity 96

region 166..281

id N70479

est

- (ix) FEATURE:
 - (A) NAME/KEY: other
 - (B) LOCATION: complement (380..486)
 - (C) IDENTIFICATION METHOD: blastn
 - (D) OTHER INFORMATION: identity 97

region 54..160

id N70479

est

- (ix) FEATURE:
 - (A) NAME/KEY: other
 - (B) LOCATION: complement(110..145)
 - (C) IDENTIFICATION METHOD: blastn
 - (D) OTHER INFORMATION: identity 94

region 403..438

id N70479

est

- (ix) FEATURE:
 - (A) NAME/KEY: other
 - (B) LOCATION: complement(196..229)
 - (C) IDENTIFICATION METHOD: blastn
 - (D) OTHER INFORMATION: identity 94

region 315..348

id N70479

est

- (ix) FEATURE:
 - (A) NAME/KEY: sig_peptide
 - (B) LOCATION: 90..140
 - (C) IDENTIFICATION METHOD: Von Heijne matrix
 - (D) OTHER INFORMATION: score 8.2

seg LLLITAILAVAVG/FP

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 17:

WO 99/06439 PCT/IB98/01233

7

GAGAGAAAGA ACTGACTGAR ACGTTTGAG ATG AAG AAA GTT CTC CTG ATC 113

Met Lys Lys Val Leu Leu Leu Ile -15 -10

ACA GCC ATC TTG GCA GTG GCT GTW GGT TTC CCA GTC TCT CAA GAC CAG 161

Thr Ala Ile Leu Ala Val Ala Val Gly Phe Pro Val Ser Gln Asp Gln -5 1 5

GAA CGA GAA AAA AGA AGT ATC AGT GAC AGC GAT GAA TTA GCT TCA GGR 209

Glu Arg Glu Lys Arg Ser Ile Ser Asp Ser Asp Glu Leu Ala Ser Gly
10 15 20

WTT TTT GTG TTC CCT TAC CCA TAT CCA TTT CGC CCA CTT CCA CCA ATT 257

Xaa Phe Val Phe Pro Tyr Pro Tyr Pro Phe Arg Pro Leu Pro Pro Ile 25 30 35

CCA TTT CCA AGA TTT CCA TGG TTT AGA CGT AAN TTT CCT ATT CCA ATA 305

Pro Phe Pro Arg Phe Pro Trp Phe Arg Arg Xaa Phe Pro Ile Pro Ile 40 45 50 55

CCT GAA TCT GCC CCT ACA ACT CCC CTT CCT AGC GAA AAG TAAACAARAA 354

Pro Glu Ser Ala Pro Thr Thr Pro Leu Pro Ser Glu Lys 60 65

GGAAAAGTCA CRATAAACCT GGTCACCTGA AATTGAAATT GAGCCACTTC CTTGAARAAT 414

CAAAATTCCT GTTAATAAAA RAAAAACAAA TGTAATTGAA ATAGCACACA GCATTCTCTA 474

GTCAATATCT TTAGTGATCT TCTTTAATAA ACATGAAAGC AAAAAAAAA AA 526

(2) INFORMATION FOR SEQ ID NO: 18:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 17 amino acids
 - (B) TYPE: AMINO ACID
 - (D) TOPOLOGY: LINEAR
- (ii) MOLECULE TYPE: PROTEIN
- (vi) ORIGINAL SOURCE:
 - (A) ORGANISM: Homo Sapiens
- (ix) FEATURE:
 - (A) NAME/KEY: sig_peptide
 - (B) LOCATION: 1..17
 - (C) IDENTIFICATION METHOD: Von Heijne matrix
 - (D) OTHER INFORMATION: score 8.2 seq LLLITAILAVAVG/FP
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 18:

Met Lys Lys Val Leu Leu Leu Ile Thr Ala Ile Leu Ala Val Ala Val 1 5 10 15

Gly

(2) INFORMATION FOR SEQ ID NO: 19:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 822 base pairs
 - (B) TYPE: NUCLEIC ACID
 - (C) STRANDEDNESS: DOUBLE
 - (D) TOPOLOGY: LINEAR
- (ii) MOLECULE TYPE: CDNA
- (vi) ORIGINAL SOURCE:
 - (A) ORGANISM: Homo Sapiens
 - (D) DEVELOPMENTAL STAGE: Fetal
 - (F) TISSUE TYPE: kidney
- (ix) FEATURE:
 - (A) NAME/KEY: other
 - (B) LOCATION: 260..464
 - (C) IDENTIFICATION METHOD: blastn
 - (D) OTHER INFORMATION: identity 96 region 153..357

id H57434

est

- (ix) FEATURE:
 - (A) NAME/KEY: other
 - (B) LOCATION: 118..184
 - (C) IDENTIFICATION METHOD: blastn
 - (D) OTHER INFORMATION: identity 98 region 98..164

region 98..164 id H57434

est

- ·(ix) FEATURE:
 - (A) NAME/KEY: other
 - (B) LOCATION: 56..113
 - (C) IDENTIFICATION METHOD: blastn
 - (D) OTHER INFORMATION: identity 98

region 35..92

id H57434

est

- (ix) FEATURE:
 - (A) NAME/KEY: other
 - (B) LOCATION: 454..485
 - (C) IDENTIFICATION METHOD: blastn
 - (D) OTHER INFORMATION: identity 100

region 348..379

id H57434

est

(ix) FEATURE:

WO 99/06439 PCT/IB98/01233

9

(A) NAME/KEY: other
(B) LOCATION: 118..545

(C) IDENTIFICATION METHOD: blastn

(D) OTHER INFORMATION: identity 98

region 1..428 id N27248

est

(ix) FEATURE:

- (A) NAME/KEY: other
- (B) LOCATION: 65..369
- (C) IDENTIFICATION METHOD: blastn
- (D) OTHER INFORMATION: identity 98

region 41..345

id H94779

est

(ix) FEATURE:

- (A) NAME/KEY: other
- (B) LOCATION: 61..399
- (C) IDENTIFICATION METHOD: blastn
- (D) OTHER INFORMATION: identity 99

region 6..344

id H09880

est

(ix) FEATURE:

- (A) NAME/KEY: other
- (B) LOCATION: 408..458
- (C) IDENTIFICATION METHOD: blastn
- (D) OTHER INFORMATION: identity 92

region 355..405

id H09880

est

(ix) FEATURE:

- (A) NAME/KEY: other
- (B) LOCATION: 60..399 .
- (C) IDENTIFICATION METHOD: blastn
- (D) OTHER INFORMATION: identity 97

region 56..395

id H29351

est

(ix) FEATURE:

- (A) NAME/KEY: other
- (B) LOCATION: 393..432
- (C) IDENTIFICATION METHOD: blastn

(D) OTHER INFORMATION: identity 90

region 391..430

id H29351

est

(ix) FEATURE:

- (A) NAME/KEY: sig_peptide
- (B) LOCATION: 346..408
- (C) IDENTIFICATION METHOD: Von Heijne matrix
- (D) OTHER INFORMATION: score 5.5

seq SFLPSALVIWTSA/AF

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 19:

ACTCCTTTTA GCATAGGGGC TTCGGCGCCA GCGGCCAGCG CTAGTCGGTC TGGTAAGTGC 60

CTGATGCCGA GTTCCGTCTC TCGCGTCTTT TCCTGGTCCC AGGCAAAGCG GASGNAGATC 120

CTCAAACGGC CTAGTGCTTC GCGCTTCCGG AGAAAATCAG CGGTCTAATT AATTCCTCTG

GTTTGTTGAA GCAGTTACCA AGAATCTTCA ACCCTTTCCC ACAAAAGCTA ATTGAGTACA 240

CGTTCCTGTT GAGTACACGT TCCTGTTGAT TTACAAAAGG TGCAGGTATG AGCAGGTCTG 300

AAGACTAACA TTTTGTGAAG TTGTAAAACA GAAAACCTGT TAGAA ATG TGG TGT TTT 357

Met Trp Trp Phe -20

CAG CAA GGC CTC AGT TTC CTT CCT TCA GCC CTT GTA ATT TGG ACA TCT 405

Gln Gln Gly Leu Ser Phe Leu Pro Ser Ala Leu Val Ile Trp Thr Ser
-15 -10 -5

GCT GCT TTC ATA TTT TCA TAC ATT ACT GCA GTA ACA CTC CAC CAT ATA 453

Ala Ala Phe Ile Phe Ser Tyr Ile Thr Ala Val Thr Leu His His Ile 1 5 10 15

GAC CCG GCT TTA CCT TAT ATC AGT GAC ACT GGT ACA GTA GCT CCA RAA

Asp Pro Ala Leu Pro Tyr Ile Ser Asp Thr Gly Thr Val Ala Pro Xaa 20 25 30

AAA TGC TTA TTT GGG GCA ATG CTA AAT ATT GCG GCA GTT TTA TGT CAA 549

Lys Cys Leu Phe Gly Ala Met Leu Asn Ile Ala Ala Val Leu Cys Gln 35 40 45

AAA TAGAAATCAG GAARATAATT CAACTTAAAG AAKTTCATTT CATGACCAAA 602

Lys

CTCTTCARAA ACATGTCTTT ACAAGCATAT CTCTTGTATT GCTTTCTACA CTGTTGAATT 662

GTCTGGCAAT ATTTCTGCAG TGGAAAATTT GATTTARMTA GTTCTTGACT GATAAATATG 722

GTAAGGTGGG CTTTTCCCCC TGTGTAATTG GCTACTATGT CTTACTGAGC CAAGTTGTAW 782

TTTGAAATAA AATGATATGA GAGTGACACA AAAAAAAAA 822

(2) INFORMATION FOR SEQ ID NO: 20:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 21 amino acids
- (B) TYPE: AMINO ACID
- (D) TOPOLOGY: LINEAR
- (ii) MOLECULE TYPE: PROTEIN
- (vi) ORIGINAL SOURCE:
 - (A) ORGANISM: Homo Sapiens
- (ix) FEATURE:
 - (A) NAME/KEY: sig_peptide
 - (B) LOCATION: 1..21
 - (C) IDENTIFICATION METHOD: Von Heijne matrix
 - (D) OTHER INFORMATION: score 5.5

seq SFLPSALVIWTSA/AF

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 20:

Met Trp Trp Phe Gln Gln Gly Leu Ser Phe Leu Pro Ser Ala Leu Val

Ile Trp Thr Ser Ala 20

- (2) INFORMATION FOR SEQ ID NO: 21:
 - (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 405 base pairs
 - (B) TYPE: NUCLEIC ACID
 - (C) STRANDEDNESS: DOUBLE
 - (D) TOPOLOGY: LINEAR
 - (ii) MOLECULE TYPE: CDNA
 - (vi) ORIGINAL SOURCE:
 - (A) ORGANISM: Homo Sapiens
 - (F) TISSUE TYPE: Testis
 - (ix) FEATURE:
 - (A) NAME/KEY: other
 - (B) LOCATION: complement(103..398)
 - (C) IDENTIFICATION METHOD: blastn
 - (D) OTHER INFORMATION: identity 96

region 1..296

id AA442893

est

- (ix) FEATURE:
 - (A) NAME/KEY: sig_peptide
 - (B) LOCATION: 185..295
 - (C) IDENTIFICATION METHOD: Von Heijne matrix
 - (D) OTHER INFORMATION: score 5.9

seq LSYASSALSPCLT/AP

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 21:

ATCACCTTCT TCTCCATCCT TSTCTGGGCC AGTCCCCARC CCAGTCCCTC TCCTGACCTG 60

CCCAGCCCAA GTCAGCCTTC AGCACGCGCT TTTCTGCACA CAGATATTCC AGGCCTACCT 120

GGCATTCCAG GACCTCCGMA ATGATGCTCC AGTCCCTTAC AAGCGCTTCC TGGATGAGGG

TGGC ATG GTG CTG ACC ACC CTC CCC TTG CCC TCT GCC AAC AGC CCT GTG 229

Met Val Leu Thr Thr Leu Pro Leu Pro Ser Ala Asn Ser Pro Val -35 -30 -25

AAC ATG CCC ACC ACT GGC CCC AAC AGC CTG AGT TAT GCT AGC TCT GCC 277

Asn Met Pro Thr Thr Gly Pro Asn Ser Leu Ser Tyr Ala Ser Ser Ala -20 -15 -10

CTG TCC CCC TGT CTG ACC GCT CCA AAK TCC CCC CGG CTT GCT ATG ATG 325

Leu Ser Pro Cys Leu Thr Ala Pro Xaa Ser Pro Arg Leu Ala Met Met
-5 1 5 10

CCT GAC AAC TAAATATCCT TATCCAAATC AATAAARWRA RAATCCTCCC TCCARAAGGG 384 Pro Asp Asn

TTTCTAAAAA CAAAAAAAA A

- (2) INFORMATION FOR SEQ ID NO: 22:
 - (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 37 amino acids
 - (B) TYPE: AMINO ACID
 - (D) TOPOLOGY: LINEAR
 - (ii) MOLECULE TYPE: PROTEIN
 - (vi) ORIGINAL SOURCE:
 - (A) ORGANISM: Homo Sapiens
 - (ix) FEATURE:
 - (A) NAME/KEY: sig_peptide
 - (B) LOCATION: 1..37
 - (C) IDENTIFICATION METHOD: Von Heijne matrix
 - (D) OTHER INFORMATION: score 5.9

seq LSYASSALSPCLT/AP

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 22:

Met Val Leu Thr Thr Leu Pro Leu Pro Ser Ala Asn Ser Pro Val Asn 1 5 10 15

Met Pro Thr Thr Gly Pro Asn Ser Leu Ser Tyr Ala Ser Ser Ala Leu 20 25 30

Ser Pro Cys Leu Thr 35 *

- (2) INFORMATION FOR SEQ ID NO: 23:
 - (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 496 base pairs
 - (B) TYPE: NUCLEIC ACID
 - (C) STRANDEDNESS: DOUBLE
 - (D) TOPOLOGY: LINEAR
 - (ii) MOLECULE TYPE: CDNA
 - (vi) ORIGINAL SOURCE:
 - (A) ORGANISM: Homo Sapiens
 - (F) TISSUE TYPE: Cancerous prostate
 - (ix) FEATURE:
 - (A) NAME/KEY: other
 - (B) LOCATION: 149..331
 - (C) IDENTIFICATION METHOD: blastn
 - (D) OTHER INFORMATION: identity 98 region 1..183 id AA397994

est

- (ix) FEATURE:
 - (A) NAME/KEY: other
 - (B) LOCATION: 328..485
 - (C) IDENTIFICATION METHOD: blastn
 - (D) OTHER INFORMATION: identity 96 region 179..336 id AA397994

est

- (ix) FEATURE:
 - (A) NAME/KEY: other
 - (B) LOCATION: complement(182..496)
 - (C) IDENTIFICATION METHOD: blastn
 - (D) OTHER INFORMATION: identity 97 region 14..328

id AA399680

est

- (ix) FEATURE:
 - (A) NAME/KEY: sig_peptide
 - (B) LOCATION: 196..240
 - (C) IDENTIFICATION METHOD: Von Heijne matrix
 - (D) OTHER INFORMATION: score 5.5

seq ILSTVTALTFAXA/LD

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 23:

AAAAAATTGG TCCCAGTTTT CACCCTGCCG CAGGGCTGGC TGGGGAGGGC AGCGGTTTAG

PCT/IB98/01233

ATTAGCCGTG GCCTAGGCCG TTTAACGGGG TGACACGAGC NTGCAGGGCC GAGTCCAAGG 120

CCCGGAGATA GGACCAACCG TCAGGAATGC GAGGAATGTT TTTCTTCGGA CTCTATCGAG

GCACACAGAC AGACC ATG GGG ATT CTG TCT ACA GTG ACA GCC TTA ACA TTT 231

Met Gly Ile Leu Ser Thr Val Thr Ala Leu Thr Phe
-15 -10 -5

GCC ARA GCC CTG GAC GGC TGC AGA AAT GGC ATT GCC CAC CCT GCA AGT 279

Ala Xaa Ala Leu Asp Gly Cys Arg Asn Gly Ile Ala His Pro Ala Ser l 5 $$ 10

GAG AAG CAC AGA CTC GAG AAA TGT AGG GAA CTC GAG ASC ASC CAC TCG 327

Glu Lys His Arg Leu Glu Lys Cys Arg Glu Leu Glu Xaa Xaa His Ser 15 20 25

GCC CCA GGA TCA ACC CAS CAC CGA AGA AAA ACA ACC AGA AGA AAT TAT 375

Ala Pro Gly Ser Thr Xaa His Arg Arg Lys Thr Thr Arg Arg Asn Tyr 30 35 40 45

TCT TCA GCC TGAAATGAAK CCGGGATCAA ATGGTTGCTG ATCARAGCCC ATATTTAAAT 434 Ser Ser Ala

TGGAAAAGTC AAATTGASCA TTATTAAATA AAGCTTGTTT AATATGTCTC AAACAAAAAA 494

AA 496

(2) INFORMATION FOR SEQ ID NO: 24:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 15 amino acids
 - (B) TYPE: AMINO ACID
 - (D) TOPOLOGY: LINEAR
- (ii) MOLECULE TYPE: PROTEIN
- (vi) ORIGINAL SOURCE:
 - (A) ORGANISM: Homo Sapiens
- (ix) FEATURE:
 - (A) NAME/KEY: sig_peptide
 - (B) LOCATION: 1..15
 - (C) IDENTIFICATION METHOD: Von Heijne matrix
 - (D) OTHER INFORMATION: score 5.5

seq ILSTVTALTFAXA/LD

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 24:

Met Gly Ile Leu Ser Thr Val Thr Ala Leu Thr Phe Ala Xaa Ala 1 5 10 15

(2) INFORMATION FOR SEQ ID NO: 25:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 623 base pairs
 - (B) TYPE: NUCLEIC ACID
 - (C) STRANDEDNESS: DOUBLE
 - (D) TOPOLOGY: LINEAR
- (ii) MOLECULE TYPE: CDNA
- (vi) ORIGINAL SOURCE:
 - (A) ORGANISM: Homo Sapiens
 - (F) TISSUE TYPE: Testis
- (ix) FEATURE:
 - (A) NAME/KEY: sig_peptide
 - (B) LOCATION: 49..96
 - (C) IDENTIFICATION METHOD: Von Heijne matrix
 - (D) OTHER INFORMATION: score 10.1

seq LVLTLCTLPLAVA/SA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 25:

AAAGATCCCT GCAGCCCGGC AGGAGAGAAG GCTGAGCCTT CTGGCGTC ATG GAG AGG 57

Met Glu Arg -15

CTC GTC CTA ACC CTG TGC ACC CTC CCG CTG GCT GTG GCG TCT GCT GGC 105

Leu Val Leu Thr Leu Cys Thr Leu Pro Leu Ala Val Ala Ser Ala Gly
-10 -5 1

TGC GCC ACG ACG CCA GCT CGC AAC CTG AGC TGC TAC CAG TGC TTC AAG 153

Cys Ala Thr Thr Pro Ala Arg Asn Leu Ser Cys Tyr Gln Cys Phe Lys 5 10 15

GTC AGC AGC TGG ACG GAG TGC CCG CCC ACC TGG TGC AGC CCG CTG GAC

Val Ser Ser Trp Thr Glu Cys Pro Pro Thr Trp Cys Ser Pro Leu Asp 20 25 30 35

CAA GTC TGC ATC TCC AAC GAG GTG GTC GTC TCT TTT AAA TGG AGT GTA 249

Gln Val Cys Ile Ser Asn Glu Val Val Val Ser Phe Lys Trp Ser Val
40 45 50

CGC GTC CTG CTC AGC AAA CGC TGT GCT CCC AGA TGT CCC AAC GAC AAC 297

Arg Val Leu Leu Ser Lys Arg Cys Ala Pro Arg Cys Pro Asn Asp Asn 55 60 65

ATG AAK TTC GAA TGG TCG CCG GCC CCC ATG GTG CAA GGC GTG ATC ACC

Met Xaa Phe Glu Trp Ser Pro Ala Pro Met Val Gln Gly Val Ile Thr 70 75 80

AGG CGC TGC TGT TCC TGG GCT CTC TGC AAC AGG GCA CTG ACC CCA CAG

Arg Arg Cys Cys Ser Trp Ala Leu Cys Asn Arg Ala Leu Thr Pro Gln 85 90 95

GAG GGG CGC TGG GCC CTG CRA GGG GGG CTC CTG CTC CAG GAC CCT TCG

Glu Gly Arg Trp Ala Leu Xaa Gly Gly Leu Leu Gln Asp Pro Ser 100 105 110 115

AGG GGC ARA AAA ACC TGG GTG CGG CCA CAG CTG GGG CTC CCA CTC TGC 489

Arg Gly Xaa Lys Thr Trp Val Arg Pro Gln Leu Gly Leu Pro Leu Cys 120 125 130

CTT CCC AWT TCC AAC CCC CTC TGC CCA RGG GAA ACC CAG GAA GGA 534

Leu Pro Xaa Ser Asn Pro Leu Cys Pro Xaa Glu Thr Gln Glu Gly 135 140 145

TAACACTGTG GGTGCCCCCA CCTGTGCATT GGGACCACRA CTTCACCCTC TTGGARACAA 594

TAAACTCTCA TGCCCCCAAA AAAAAAAAA 623

(2) INFORMATION FOR SEQ ID NO: 26:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 16 amino acids
 - (B) TYPE: AMINO ACID
 - (D) TOPOLOGY: LINEAR
- (ii) MOLECULE TYPE: PROTEIN
- (vi) ORIGINAL SOURCE:
 - (A) ORGANISM: Homo Sapiens
- (ix) FEATURE:
 - (A) NAME/KEY: sig_peptide
 - (B) LOCATION: $1..\overline{16}$
 - (C) IDENTIFICATION METHOD: Von Heijne matrix
 - (D) OTHER INFORMATION: score 10.1

seq LVLTLCTLPLAVA/SA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 26:

Met Glu Arg Leu Val Leu Thr Leu Cys Thr Leu Pro Leu Ala Val Ala 1 5 10 15

(2) INFORMATION FOR SEQ ID NO: 27:

WO 99/06439

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 848 base pairs
- (B) TYPE: NUCLEIC ACID
- (C) STRANDEDNESS: DOUBLE
- (D) TOPOLOGY: LINEAR
- (ii) MOLECULE TYPE: CDNA
- (vi) ORIGINAL SOURCE:
 - (A) ORGANISM: Homo Sapiens
 - (D) DEVELOPMENTAL STAGE: Fetal
 - (F) TISSUE TYPE: kidney
- (ix) FEATURE:
 - (A) NAME/KEY: sig_peptide
 - (B) LOCATION: 32..73
 - (C) IDENTIFICATION METHOD: Von Heijne matrix
 - (D) OTHER INFORMATION: score 10.7

seq LWLLFFLVTAIHA/EL

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 27:

AACTTTGCCT TGTGTTTTCC ACCCTGAAAG A ATG TTG TGG CTG CTC TTT TTT CTG 55

Met Leu Trp Leu Leu Phe Phe Leu

PCT/IB98/01233

GTG ACT GCC ATT CAT GCT GAA CTC TGT CAA CCA GGT GCA GAA AAT GCT

Val Thr Ala Ile His Ala Glu Leu Cys Gln Pro Gly Ala Glu Asn Ala
-5 1 5 10

TTT AAA GTG AGA CTT AGT ATC AGA ACA GCT CTG GGA GAT AAA GCA TAT

Phe Lys Val Arg Leu Ser Ile Arg Thr Ala Leu Gly Asp Lys Ala Tyr 15 20 25

GCC TGG GAT ACC AAT GAA GAA TAC CTC TTC AAA GCG ATG GTA GCT TTC 199

Ala Trp Asp Thr Asn Glu Glu Tyr Leu Phe Lys Ala Met Val Ala Phe 30 35 40

TCC ATG AGA AAA GTT CCC AAC AGA GAA GCA ACA GAA ATT TCC CAT GTC 247

Ser Met Arg Lys Val Pro Asn Arg Glu Ala Thr Glu Ile Ser His Val 45 50 55

CTA CTT TGC AAT GTA ACC CAG AGG GTA TCA TTC TGG TTT GTG GTT ACA 295

Leu Leu Cys Asn Val Thr Gln Arg Val Ser Phe Trp Phe Val Val Thr 60 65 70

GAC CCT TCA AAR AAT CAC ACC CTT CCT GCT GTT GAG GTG CAA TCA GCC 343

Asp Pro Ser Lys Asn His Thr Leu Pro Ala Val Glu Val Gln Ser Ala 75 80 85 90

ATA AGA ATG AAC AAG AAC CGG ATC AAC AAT GCC TTC TTT CTA AAT GAC 39!

Ile Arg Met Asn Lys Asn Arg Ile Asn Asn Ala Phe Phe Leu Asn Asp

95 100 105

CAA ACT CTG GAA TTT TTA AAA ATC CCT TCC ACA CTT GCA CCA CCC ATG 439

Gln Thr Leu Glu Phe Leu Lys Ile Pro Ser Thr Leu Ala Pro Pro Met 110 115 120

GAC CCA TCT GTG CCC ATC TGG ATT ATT ATA TTT GGT GTG ATA TTT TGC 487

Asp Pro Ser Val Pro Ile Trp Ile Ile Ile Phe Gly Val Ile Phe Cys 125 130 135

ATC ATC ATA GTT GCA ATT GCA CTA CTG ATT TTA TCA GGG ATC TGG CAA 535

Ile Ile Ile Val Ala Ile Ala Leu Leu Ile Leu Ser Gly Ile Trp Gln 140 145 150

CGT ADA ARA AAG AAC AAA GAA CCA TCT GAA GTG GAT GAC GCT GAA RAT 583

Arg Xaa Xaa Lys Asn Lys Glu Pro Ser Glu Val Asp Asp Ala Glu Xaa 155 160 165 170

AAK TGT GAA AAC ATG ATC ACA ATT GAA AAT GGC ATC CCC TCT GAT CCC 631

Xaa Cys Glu Asn Met Ile Thr Ile Glu Asn Gly Ile Pro Ser Asp Pro 175 180 185

CTG GAC ATG AAG GGA GGG CAT ATT AAT GAT GCC TTC ATG ACA GAG GAT 679

Leu Asp Met Lys Gly Gly His Ile Asn Asp Ala Phe Met Thr Glu Asp 190 195 200

GAG AGG CTC ACC CCT CTC TGAAGGGCTG TTGTTCTGCT TCCTCAARAA

Glu Arg Leu Thr Pro Leu 205

ATTAAACATT TGTTTCTGTG TGACTGCTGA GCATCCTGAA ATACCAAGAG CAGATCATAT 787

WTTTTGTTTC ACCATTCTTC TTTTGTAATA AATTTTGAAT GTGCTTGAAA AAAAAAAAA 847

C 848

(2) INFORMATION FOR SEQ ID NO: 28:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 14 amino acids
 - (B) TYPE: AMINO ACID
 - (D) TOPOLOGY: LINEAR
- (ii) MOLECULE TYPE: PROTEIN
- (vi) ORIGINAL SOURCE:
 - (A) CRGANISM: Homo Sapiens
- (ix) FEATURE:

WO 99/06439 PCT/IB98/01233

19

(A) NAME/KEY: sig_peptide

- (B) LOCATION: 1..14
- (C) IDENTIFICATION METHOD: Von Heijne matrix
- (D) OTHER INFORMATION: score 10.7 seq LWLLFFLVTAIHA/EL
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 28:

Met Leu Trp Leu Leu Phe Phe Leu Val Thr Ala Ile His Ala 1 5 10

- (2) INFORMATION FOR SEQ ID NO: 29:
 - (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 25 base pairs
 - (B) TYPE: NUCLEIC ACID
 - (C) STRANDEDNESS: SINGLE
 - (D) TOPOLOGY: LINEAR
 - (ii) MOLECULE TYPE: Other nucleic acid
 - (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 29:

GGGAAGATGG AGATAGTATT GCCTG

25

- (2) INFORMATION FOR SEQ ID NO: 30:
 - (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 26 base pairs
 - (B) TYPE: NUCLEIC ACID
 - (C) STRANDEDNESS: SINGLE
 - (D) TOPOLOGY: LINEAR
 - (ii) MOLECULE TYPE: Other nucleic acid
 - (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 30:

CTGCCATGTA CATGATAGAG AGATTC

26

- (2) INFORMATION FOR SEQ ID NO: 31:
 - (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 546 base pairs
 - (B) TYPE: NUCLEIC ACID
 - (C) STRANDEDNESS: DOUBLE
 - (D) TOPOLOGY: LINEAR
 - (ii) MOLECULE TYPE: Genomic DNA
 - (ix) FEATURE:
 - (A) NAME/KEY: promoter
 - (B) LOCATION: 1..517

(ix) FEATURE:

(A) NAME/KEY: transcription start site

(B) LOCATION: 518

(ix) FEATURE:

(A) NAME/KEY: TF binding-site

(B) LOCATION: 17..25

(C) IDENTIFICATION METHOD: matinspector prediction

(D) OTHER INFORMATION: name CMYB_01 score 0.983

sequence TGTCAGTTG

(ix) FEATURE:

(A) NAME/KEY: TF binding-site

(B) LOCATION: complement(18..27)

(C) IDENTIFICATION METHOD: matinspector prediction

(D) OTHER INFORMATION: name MYOD_Q6 score 0.961

sequence CCCAACTGAC

(ix) FEATURE:

(A) NAME/KEY: TF binding-site

(B) LOCATION: complement (75..85)

(C) IDENTIFICATION METHOD: matinspector prediction

(D) OTHER INFORMATION: name S8_01 score 0.960

sequence AATAGAATTAG

(ix) FEATURE:

(A) NAME/KEY: TF binding-site

(B) LOCATION: 94..104

(C) IDENTIFICATION METHOD: matinspector prediction

(D) OTHER INFORMATION: name S8_01
score 0.966
sequence AACTAAATTAG

(ix) FEATURE:

(A) NAME/KEY: TF binding-site

(B) LOCATION: complement(129..139)

(C) IDENTIFICATION METHOD: matinspector prediction

(D) OTHER INFORMATION: name DELTAEF1_01 score 0.960 sequence GCACACCTCAG

(ix) FEATURE:

(A) NAME/KEY: TF binding-site

(B) LOCATION: complement(155..165)

(C) IDENTIFICATION METHOD: matinspector prediction

(D) OTHER INFORMATION: name GATA C score 0.964

sequence AGATAAATCCA

(ix) FEATURE:

(A) NAME/KEY: TF binding-site

(B) LOCATION: 170..178

(C) IDENTIFICATION METHOD: matinspector prediction

(D) OTHER INFORMATION: name CMYB_01
score 0.958
sequence CTTCAGTTG

(ix) FEATURE:

- (A) NAME/KEY: TF binding-site
- (B) LOCATION: 176..189
- (C) IDENTIFICATION METHOD: matinspector prediction
- (D) OTHER INFORMATION: name GATA1 02 score 0.959

sequence TTGTAGATAGGACA

(ix) FEATURE:

- (A) NAME/KEY: TF binding-site
- (B) LOCATION: 180..190
- (C) IDENTIFICATION METHOD: matinspector prediction
- (D) OTHER INFORMATION: name GATA C score $0.9\overline{5}$ 3 sequence AGATAGGACAT

(ix) FEATURE:

- (A) NAME/KEY: TF binding-site
- (B) LOCATION: 284..299
- (C) IDENTIFICATION METHOD: matinspector prediction
- (D) OTHER INFORMATION: name TAL1ALPHAE47 01

score 0.973

sequence CATAACAGATGGTAAG

(ix) FEATURE:

- (A) NAME/KEY: TF binding-site
- (B) LOCATION: 284..299
- (C) IDENTIFICATION METHOD: matinspector prediction
- (D) OTHER INFORMATION: name TAL1BETAE47 01 score 0.983

sequence CATAACAGATGGTAAG

(ix) FEATURE:

- (A) NAME/KEY: TF binding-site
- (B) LOCATION: 284..299
- (C) IDENTIFICATION METHOD: matinspector prediction
- (D) OTHER INFORMATION: name TAL1BETAITF2_01 score 0.978

sequence CATAACAGATGGTAAG

(ix) FEATURE:

- (A) NAME/KEY: TF binding-site
- (B) LOCATION: complement(287..296)
- (C) IDENTIFICATION METHOD: matinspector prediction
- (D) OTHER INFORMATION: name MYOD 06 score 0.954 sequence ACCATCTGTT

(ix) FEATURE:

- (A) NAME/KEY: TF binding-site
- (B) LOCATION: complement(302..314)
- (C) IDENTIFICATION METHOD: matinspector prediction
- (D) OTHER INFORMATION: name GATA1 04 score 0.953 sequence TCAAGATAAAGTA

(ix) FEATURE:

- (A) NAME/KEY: TF binding-site
- (B) LOCATION: 393..405
- (C) IDENTIFICATION METHOD: matinspector prediction
- (D) OTHER INFORMATION: name IK1_01

score 0.963 sequence AGTTGGGAATTCC

		RE	U	T	Α	E	F	x)	i	(
--	--	----	---	---	---	---	---	----	---	---

- (A) NAME/KEY: TF binding-site
- (B) LOCATION: 393..404
- (C) IDENTIFICATION METHOD: matinspector prediction
- (D) OTHER INFORMATION: name IK2 01 score $0.\overline{9}85$

sequence AGTTGGGAATTC

(ix) FEATURE:

- (A) NAME/KEY: TF binding-site
- (B) LOCATION: 396..405
- (C) IDENTIFICATION METHOD: matinspector prediction
- (D) OTHER INFORMATION: name CREL 01 score 0.962

sequence TGGGAATTCC

(ix) FEATURE:

- (A) NAME/KEY: TF binding-site
- (B) LOCATION: 423..436
- (C) IDENTIFICATION METHOD: matinspector prediction
- (D) OTHER INFORMATION: name GATA1 02 score 0.950 sequence TCAGTGATATGGCA

(ix) FEATURE:

- (A) NAME/KEY: TF binding-site
- (B) LOCATION: complement (478..489)
- (C) IDENTIFICATION METHOD: matinspector prediction
- (D) OTHER INFORMATION: name SRY 02 score $0.\overline{9}51$

sequence TAAAACAAAACA

(ix) FEATURE:

- (A) NAME/KEY: TF binding-site
- (B) LOCATION: 486..493
- (C) IDENTIFICATION METHOD: matinspector prediction
- (D) OTHER INFORMATION: name E2F 02 score $0.\overline{9}57$ sequence TTTAGCGC

(ix) FEATURE:

- (A) NAME/KEY: TF binding-site
- (B) LOCATION: complement (514..521)
- (C) IDENTIFICATION METHOD: matinspector prediction
- (D) OTHER INFORMATION: name MZF1 01 score 0.975 sequence TGAGGGGA
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 31:

TGAGTGCAGT GTTACATGTC AGTTGGGTTA AGTTTGTTAA TGTCATTCAA ATCTTCTATG 60 TCTTGATTTG CCTGCTAATT CTATTATTTC TGGAACTAAA TTAGTTTGAT GGTTCTATTA 120 GTTATTGACT GAGGTGTGCT AATCTCCCAT TATGTGGATT TATCTATTTC TTCAGTTGTA GATAGGACAT TGATAGATAC ATAAGTACCA GGACAAAAGC AGGGAGATCT TTTTTCCAAA 240

WO 99/06439	PCT/IB98/01233
W () 33/00433	PC17IB98

. 23	
ATCAGGAGAA AAAAATGACA TCTGGAAAAC CTATAGGGAA AGGCATAACA GATGGTAAGG	300
ATACTTTATC TTGAGTAGGA GAGCCTTCCT GTGGCAACGT GGAGAAGGGA AGAGGTCGTA	360
GAATTGAGGA GTCAGCTCAG TTAGAAGCAG GGAGTTGGGA ATTCCGTTCA TGTGATTTAG	420
CATCAGTGAT ATGGCAAATG TGGGACTAAG GGTAGTGATC AGAGGGTTAA AATTGTGTGT	480
TTTGTTTTAG CGCTGCTGGG GCATCGCCTT GGGTCCCCTC AAACAGATTC CCATGAATCT	540
CTTCAT	546
(2) INFORMATION FOR SEQ ID NO: 32: (i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 23 base pairs (B) TYPE: NUCLEIC ACID (C) STRANDEDNESS: SINGLE (D) TOPOLOGY: LINEAR (ii) MOLECULE TYPE: Other nucleic acid (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 32:	
GTACCAGGGA CTGTGACCAT TGC	23
(2) INFORMATION FOR SEQ ID NO: 33: (i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 24 base pairs (B) TYPE: NUCLEIC ACID (C) STRANDEDNESS: SINGLE (D) TOPOLOGY: LINEAR (ii) MOLECULE TYPE: Other nucleic acid (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 33:	
CTGTGACCAT TGCTCCCAAG AGAG	24
(2) INFORMATION FOR SEQ ID NO: 34: (i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 861 base pairs (B) TYPE: NUCLEIC ACID (C) STRANDEDNESS: DOUBLE (D) TOPOLOGY: LINEAR (ii) MOLECULE TYPE: Genomic DNA	

(iz) FEATURE:

(A) NAME/KEY: promoter

(B) LOCATION: 1..806

(ix) FEATURE:

(A) NAME/KEY: transcription start site

(B) LOCATION: 807

(ix) FEATURE:

(A) NAME/KEY: TF binding-site

(B) LOCATION: complement(60..70)

(C) IDENTIFICATION METHOD: matinspector prediction

(D) OTHER INFORMATION: name NFY_Q6

score $0.\overline{9}56$

sequence GGACCAATCAT

(ix) FEATURE:

(A) NAME/KEY: TF binding-site

(B) LOCATION: 70..77

(C) IDENTIFICATION METHOD: matinspector prediction

(D) OTHER INFORMATION: name MZF1_01 score 0.962

sequence CCTGGGGA

(ix) FEATURE:

(A) NAME/KEY: TF binding-site

(B) LOCATION: 124..132

(C) IDENTIFICATION METHOD: matinspector prediction

(D) OTHER INFORMATION: name CMYB_01

score 0.994

sequence TGACCGTTG

(ix) FEATURE:

(A) NAME/KEY: TF binding-site

(B) LOCATION: complement(126..134)

(C) IDENTIFICATION METHOD: matinspector prediction

(D) OTHER INFORMATION: name VMYB 02

score 0.985

sequence TCCAACGGT

(ix) FEATURE:

(A) NAME/KEY: TF binding-site

(B) LOCATION: 135..143

(C) IDENTIFICATION METHOD: matinspector prediction

(D) OTHER INFORMATION: name STAT_01

score 0.968

sequence TTCCTGGAA

(ix) FEATURE:

(A) NAME/KEY: TF binding-site

(B) LOCATION: complement (135..143)

(C) IDENTIFICATION METHOD: matinspector prediction

(D) OTHER INFORMATION: name STAT_01

score $0.9\overline{51}$

sequence TTCCAGGAA

(ix) FEATURE:

(A) NAME/KEY: TF binding-site

(B) LOCATION: complement (252..259)

(C) IDENTIFICATION METHOD: matinspector prediction

(D) OTHER INFORMATION: name MZF1_01 score 0.956

sequence TTGGGGGA

(ix) FEATURE:

(A) NAME/KEY: TF binding-site

(B) LOCATION: 357..368

(C) IDENTIFICATION METHOD: matinspector prediction

(D) OTHER INFORMATION: name IK2_01 score 0.965

sequence GAATGGGATTTC

(ix) FEATURE:

(A) NAME/KEY: TF binding-site

(B) LOCATION: 384..391

(C) IDENTIFICATION METHOD: matinspector prediction

(D) OTHER INFORMATION: name MZF1_01 score 0.986 sequence AGAGGGGA

(ix) FEATURE:

(A) NAME/KEY: TF binding-site

- (B) LOCATION: complement (410..421)
- (C) IDENTIFICATION METHOD: matinspector prediction
- (D) OTHER INFORMATION: name SRY_02
 score 0.955
 sequence GAAAACAAAACA

(ix) FEATURE:

- (A) NAME/KEY: TF binding-site
- (B) LOCATION: 592..599
- (C) IDENTIFICATION METHOD: matinspector prediction
- (D) OTHER INFORMATION: name MZF1_01 score 0.960 sequence GAAGGGGA

(ix) FEATURE:

- (A) NAME/KEY: TF binding-site
- (B) LOCATION: 618..627
- (C) IDENTIFICATION METHOD: matinspector prediction
- (D) OTHER INFORMATION: name MYOD_Q6 score 0.981 sequence AGCATCTGCC

(ix) FEATURE:

- (A) NAME/KEY: TF binding-site
- (B) LOCATION: 632..642
- (C) IDENTIFICATION METHOD: matinspector prediction
- (D) OTHER INFORMATION: name DELTAEF1_01 score 0.953 sequence TCCCACCTTCC

(ix) FEATURE:

- (A) NAME/KEY: TF binding-site
- (B) LOCATION: complement(813..823)
- (C) IDENTIFICATION METHOD: matinspector prediction
- (D) OTHER INFORMATION: name S8_01 score 0.992 sequence GAGGCAATTAT

(ix) FEATURE:

- (A) NAME/KEY: TF binding-site
- (B) LCCATION: complement(824..831)
- (C) IDENTIFICATION METHOD: matinspector prediction

WO 99/06439 PCT/IB98/01233

26

(D) OTHER INFORMATION: name MZF1_01 score 0.986 sequence AGAGGGGA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 34:

TACTATAGGG	CACGCGTGGT	CGACGGCCGG	GCTGTTCTGG	AGCAGAGGGC	ATGTCAGTAA	60
TGATTGGTCC	CTGGGGAAGG	TCTGGCTGGC	TCCAGCACAG	TGAGGCATTT	AGGTATCTCT	120
CGGTGACCGT	TGGATTCCTG	GAAGCAGTAG	CTGTTCTGTT	TGGATCTGGT	AGGGACAGGG	180
CTCAGAGGGC	TAGGCACGAG	GGAAGGTCAG	AGGAGAAGGS	AGGSARGGCC	CAGTGAGARG	240
GGAGCATGCC	TTCCCCCAAC	CCTGGCTTSC	YCTTGGYMAM	AGGGCGKTTY	TGGGMACTTR	300
AAYTCAGGGC	CCAASCAGAA	SCACAGGCCC	AKTCNTGGCT	SMAAGCACAA	TAGCCTGAAT	360
GGGATTTCAG	GTTAGNCAGG	GTGAGAGGGG	AGGCTCTCTG	GCTTAGTTTT	GTTTTGTTTT	420
CCAAATCAAG	GTAACTTGCT	CCCTTCTGCT	ACGGGCCTTG	GTCTTGGCTT	GTCCTCACCC	480
AGTCGGAACT	CCCTACÇACT	TTCAGGAGAG	TGGTTTTAGG	CCCGTGGGGC	TGTTCTGTTC	540
CAAGCAGTGT	GAGAACATGG	CTGGTAGAGG	CTCTAGCTGT	GTGCGGGGCC	TGAAGGGGAG	600
TGGGTTCTCG	CCCAAAGAGC	ATCTGCCCAT	TTCCCACCTT	CCCTTCTCCC	ACCAGAAGCT	660
TGCCTGAGCT	GTTTGGACAA	AAATCCAAAC	CCCACTTGGC	TACTCTGGCC	TGGCTTCAGC	720
TTGGAACCCA	ATACCTAGGC	TTACAGGCCA	TCCTGAGCCA	GGGGCCTCTG	GAAATTCTCT	780
TCCTGATGGT	CCTTTAGGTT	TGGGCACAAA	ATATAATTGC	стстсссстс	TCCCATTTTC	840
TCTCTTGGGA	GCAATGGTCA	С				861

(2) INFORMATION FOR SEQ ID NO: 35:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 20 base pairs
 - (B) TYPE: NUCLEIC ACID
 - (C) STRANDEDNESS: SINGLE
 - (D) TOPOLOGY: LINEAR
- (ii) MOLECULE TYPE: Other nucleic acid
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 35:

CTGGGATGGA AGGCACGGTA

20

(2) INFORMATION FOR SEQ ID NO: 36:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 20 base pairs
 - (B) TYPE: NUCLEIC ACID

- (C) STRANDEDNESS: SINGLE
 (D) TOPOLOGY: LINEAR
- (ii) MOLECULE TYPE: Other nucleic acid
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 36:

GAGACCACAC AGCTAGACAA

20

(2) INFORMATION FOR SEQ ID NO: 37:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 555 base pairs
 - (B) TYPE: NUCLEIC ACID
 - (C) STRANDEDNESS: DOUBLE
 - (D) TOPOLOGY: LINEAR
- (ii) MOLECULE TYPE: Genomic DNA
- (ix) FEATURE:
 - (A) NAME/KEY: promoter
 - (B) LOCATION: 1..500
- (ix) FEATURE:
 - (A) NAME/KEY: transcription start site
 - (B) LOCATION: 501
- (ix) FEATURE:
 - (A) NAME/KEY: TF binding-site
 - (B) LOCATION: 191..206
 - (C) IDENTIFICATION METHOD: matinspector prediction
 - (D) OTHER INFORMATION: name ARNT_01

score 0.964

sequence GGACTCACGTGCTGCT

- (ix) FEATURE:
 - (A) NAME/KEY: TF binding-site
 - (B) LOCATION: 193..204
 - (C) IDENTIFICATION METHOD: matinspector prediction
 - (D) OTHER INFORMATION: name NMYC_01

score 0.965

sequence ACTCACGTGCTG

- (ix) FEATURE:
 - (A) NAME/KEY: TF binding-site
 - (B) LOCATION: 193..204
 - (C) IDENTIFICATION METHOD: matinspector prediction
 - (D) OTHER INFORMATION: name USF_01

score 0.985

sequence ACTCACGTGCTG

- (ix) FEATURE:
 - (A) NAME/KEY: TF binding-site
 - (2) LOCATION: complement(193..204)
 - (C) IDENTIFICATION METHOD: matinspector prediction
 - (D) OTHER INFORMATION: name USF_01

score $0.\overline{9}85$

sequence CAGCACGTGAGT

(ix) FEATURE:

- (A) NAME/KEY: TF binding-site
- (B) LOCATION: complement(193..204)
- (C) IDENTIFICATION METHOD: matinspector prediction
- (D) OTHER INFORMATION: name NMYC 01 score 0.956

sequence CAGCACGTGAGT

(ix) FEATURE:

- (A) NAME/KEY: TF binding-site
- (B) LOCATION: complement(193..204)
- (C) IDENTIFICATION METHOD: matinspector prediction
- (D) OTHER INFORMATION: name MYCMAX_02 score 0.972

sequence CAGCACGTGAGT

(ix) FEATURE:

- (A) NAME/KEY: TF binding-site
- (B) LOCATION: 195..202
- (C) IDENTIFICATION METHOD: matinspector prediction
- (D) OTHER INFORMATION: name USF_C score 0.997 sequence TCACGTGC

(ix) FEATURE:

- (A) NAME/KEY: TF binding-site
- (B) LOCATION: complement(195..202)
- (C) IDENTIFICATION METHOD: matinspector prediction
- (D) OTHER INFORMATION: name USF_C score 0.991 sequence GCACGTGA

(ix) FEATURE:

- (A) NAME/KEY: TF binding-site
- (B) LOCATION: complement(210..217)
- (C) IDENTIFICATION METHOD: matinspector prediction
- (D) OTHER INFORMATION: name MZF1_01 score 0.968 sequence CATGGGGA

(ix) FEATURE:

- (A) NAME/KEY: TF binding-site
- (B) LOCATION: 397..410
- (C) IDENTIFICATION METHOD: matinspector prediction
- (D) OTHER INFORMATION: name ELK1_02
 score 0.963
 sequence CTCTCCGGAAGCCT

(ix) FEATURE:

- (A) NAME/KEY: TF binding-site
- (B) LOCATION: 400..409
- (C) IDENTIFICATION METHOD: matinspector prediction
- (D) OTHER INFORMATION: name CETS1P54_01 score 0.974 sequence TCCGGAAGCC

(ix) FEATURE:

- (A) NAME/KEY: TF binding-site
- (B) LOCATION: complement (460..470)
- (C) IDENTIFICATION METHOD: matinspector prediction

WO 99/06439 PCT/IB98/01233

(D) OTHER INFORMATION: name AP1_Q4 score 0.963

sequence AGTGACTGAAC

(ix) FEATURE:

(A) NAME/KEY: TF binding-site

(B) LOCATION: complement (460..470)

(C) IDENTIFICATION METHOD: matinspector prediction

(D) OTHER INFORMATION: name AP1FJ Q2

score 0.961

sequence AGTGACTGAAC

(ix) FEATURE:

- (A) NAME/KEY: TF binding-site
- (B) LOCATION: 547..555
- (C) IDENTIFICATION METHOD: matinspector prediction
- (D) OTHER INFORMATION: name PADS_C score 1.000

sequence TGTGGTCTC

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 37:

CTATAGGGCA	CGCKTGGTCG	ACGGCCCGGG	CTGGTCTGGT	CTGTKGTGGA	GTCGGGTTGA	60
AGGACAGCAT	TTGTKACATC	TGGTCTACTG	CACCTTCCCT	CTGCCGTGCA	CTTGGCCTTT	120
KAWAAGCTCA	GCACCGGTGC	CCATCACAGG	GCCGGCAGCA	CACACATCCC	ATTACTCAGA	180
AGGAACTGAC	GGACTCACGT	GCTGCTCCGT	CCCCATGAGC	TCAGTGGACC	TGTCTATGTA	240
GAGCAGTCAG	ACAGTGCCTG	GGATAGAGTG	AGAGTTCAGC	CAGTAAATCC	AAGTGATTGT	300
CATTCCTGTC	TGCATTAGTA	ACTCCCAACC	TAGATGTGAA	AACTTAGTTC	TTTCTCATAG	360
GTTGCTCTGC	CCATGGTCCC	ACTGCAGACC	CAGGCACTCT	CCGGAAGCCT	GGAAATCACC	420
CGTGTCTTCT	GCCTGCTCCC	GCTCACATCC	CACACTTGTG	TTCAGTCACT	GAGTTACAGA	480
TTTTGCCTCC	TCAATTTCTC	TTGTCTTAGT	CCCATCCTCT	GTTCCCCTGG	CCAGTTTGTC	540
TAGCTGTGTG	GTCTC					555

(2) INFORMATION FOR SEQ ID NO: 38:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 352 base pairs
 - (B) TYPE: NUCLEIC ACID
 - (C) STRANDEDNESS: DOUBLE
 - (D) TOPOLOGY: LINEAR
- (ii) MOLECULE TYPE: CDNA
- (vi) ORIGINAL SOURCE:
 - (A) ORGANISM: Homo Sapiens
 - (D) DEVELOPMENTAL STAGE: Fetal
 - (F) TISSUE TYPE: liver
- (ix) FEATURE:

(A) NAME/KEY: other

(B) LOCATION: 49..196

(C) IDENTIFICATION METHOD: blastn

(D) OTHER INFORMATION: identity 97

region 12..159 id AA232452

est

(ix) FEATURE:

(A) NAME/KEY: other
(B) LOCATION: 195..337

(C) IDENTIFICATION METHOD: blastn

(D) OTHER INFORMATION: identity 96

region 157..299 id AA232452

est

(ix) FEATURE:

(A) NAME/KEY: sig_peptide

(B) LOCATION: 65..124

(C) IDENTIFICATION METHOD: Von Heijne matrix

(D) OTHER INFORMATION: score 15.8

seq LLLLLLRHGAQG/KP

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 38:

ACACATTTGC GGGAACGCAG AGCGGAGCGT NGGAGAGCGG ASRWAGCTGG ATAACAGGGG 60

ACCG ATG ATG TGG CGA CCA TCA GTT CTG CTG CTT CTG TTG CTA CTG AGG 109

Met Met Trp Arg Pro Ser Val Leu Leu Leu Leu Leu Leu Leu Arg -20 -15 -10

CAC GGG GCC CAG GGG AAG CCA TCC CCA GAC GCA GGC CCT CAT GGC CAG
His Gly Ala Gln Gly Lys Pro Ser Pro Asp Ala Gly Pro His Gly Gln
-5 1 5 10

GGG AGG GTG CAC CAG GCG GCC CCC CTG AGC GAC GCT CCC CAT GAT GAC
Gly Arg Val His Gln Ala Ala Pro Leu Ser Asp Ala Pro His Asp Asp
15
20
205

GCC CAC GGG AAC TTC CAG TAC GAC CAT GAG GCT TTC CTG GGA CGG GAA 253
Ala His Gly Asn Phe Gln Tyr Asp His Glu Ala Phe Leu Gly Arg Glu
30 35

GTG GCC AAG GAA TTC GAC CAA CTC ACC CCA GAG GAA AGC CAG GCC CGT
Val Ala Lys Glu Phe Asp Gln Leu Thr Pro Glu Glu Ser Gln Ala Arg

CTG GGG CGG ATC GTG GAC CGC ATG GAC CGC GRG GGG ACG GCA ACG GCT
Leu Gly Arg Ile Val Asp Arg Met Asp Arg Xaa Gly Thr Ala Thr Ala
60 70 75

GGG 352

(2) INFORMATION FOR SEQ ID NO: 39:

(i) SEQUENCE CHARACTERISTICS:

. 31											
(A) LENGTH: 247 base pairs(B) TYPE: NUCLEIC ACID(C) STRANDEDNESS: DOUBLE(D) TOPOLOGY: LINEAR											
(ii) MOLECULE TYPE: CDNA											
<pre>(vi) ORIGINAL SOURCE: (A) ORGANISM: Homo Sapiens (F) TISSUE TYPE: Lung (cells)</pre>											
<pre>(ix) FEATURE: (A) NAME/KEY: sig_peptide (B) LOCATION: 158241 (C) IDENTIFICATION METHOD: Von Heijne matrix (D) OTHER INFORMATION: score 11.4</pre>											
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 39:											
CTACTTTCCT GCTTACAAGA GAACTTTGGA ACCAAGGAGC AGGACTTTTA GCTGCTTGTT	60										
TTATTGCTAT TGTACCAGGC TACATATCTC GGTCAGTAGC TGGATCCTTT GATAATGAAG	120										
GCATTGCTAT TTTTGCACTT CAGTTCACAT ACTATTT ATG GGT AAA ATC TGT AAA Met Gly Lys Ile Cys Lys -25	175										
AAC TGG GTC AGT TTT TTG GAC AAT GTG CTG CTT ATC CTA TTT CTA Asn Trp Val Ser Phe Leu Asp Asn Val Leu Leu Leu Ile Leu Phe Leu -20 -15 -10	223										
TAT GGT CTC TGC TCG GGG TGG CGG Tyr Gly Leu Cys Ser Gly Trp Arg -5	247										
(2) INFORMATION FOR SEQ ID NO: 40:											
(i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 216 base pairs (B) TYPE: NUCLEIC ACID (C) STRANDEDNESS: DOUBLE (D) TOPOLOGY: LINEAR											
(ii) MOLECULE TYPE: CDNA											
<pre>(vi) ORIGINAL SOURCE: (A) ORGANISM: Homo Sapiens (F) TISSUE TYPE: Large intestine</pre>											
<pre>(ix) FEATURE: (A) NAME/KEY: sig_peptide (B) LOCATION: 61114 (C) IDENTIFICATION METHOD: Von Heijne matrix (D) OTHER INFORMATION: score 11.1</pre>											

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 40:

ATATAAAACA TCTGGAAGTT TCCAGGGGGC TGCTTTGCAT CTGAAACTGT CAGCCCCAGA	60							
ATG TTG ACA GTC GCT CTC CTA GCC CTT CTC TGT GCC TCA GCC TCT GGC Met Leu Thr Val Ala Leu Leu Ala Leu Leu Cys Ala Ser Ala Ser Gly -15 -10 -5	108							
AAT GCC ATT CAG GCC AGG TCT TCC TCC TAT AGT GGA GAG TAT GGD CTG Asn Ala Ile Gln Ala Arg Ser Ser Ser Tyr Ser Gly Glu Tyr Gly Leu 1 5 10	156							
GTG GTG GAA AGC GAT TCT CTC ATT CTG GCA ACC AGT TGG ACG GCC CCA Val Val Glu Ser Asp Ser Leu Ile Leu Ala Thr Ser Trp Thr Ala Pro 20 25 30	204							
TCA CCG CCC ACG Ser Pro Pro Thr	216							
(2) INFORMATION FOR SEQ ID NO: 41:								
(i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 202 base pairs (B) TYPE: NUCLEIC ACID (C).STRANDEDNESS: DOUBLE (D) TOPOLOGY: LINEAR								
(ii) MOLECULE TYPE: CDNA								
<pre>(vi) ORIGINAL SOURCE: (A) ORGANISM: Homo Sapiens (F) TISSUE TYPE: Pancreas</pre>								
<pre>(ix) FEATURE: (A) NAME/KEY: sig_peptide (B) LOCATION: 107145 (C) IDENTIFICATION METHOD: Von Heijne matrix (D) OTHER INFORMATION: score 10.4</pre>								
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 41:								
AAGATTCAGG CCCTCAGCAA ACAAGGAACC TGGAAAATGT AACCCTGAAT GCACGGTGGG	60							
GAGGACATGG CAAGAGAAAA GCGGCAGGAA TAAAGTGATT TTCTGA ATG GTC TTG Met Val Leu	115							
CTT CTT TGT CTA TCT TGT CTG ATT TTC TCC TGT CTG ACC TTT TCC TGG Leu Leu Cys Leu Ser Cys Leu Ile Phe Ser Cys Leu Thr Phe Ser Trp -10 -5 1 5	163							
TTA AAA ATC TGG GGG AAA ATG ACG GAC TCC AAG CCG ATG Leu Lys Ile Trp Gly Lys Met Thr Asp Ser Lys Pro Met 10 15	202							

(vi) ORIGINAL SOURCE:

(A) ORGANISM: Homo Sapiens

	(i	(A) L B) T C) S	ENGT YPE: TRAN		90 b LEIC ESS:	ase ACI DOU	pair D	s							
	(i	i) M	OLEC	ULE	TYPE	: CD	NA									
	(v		A) O	RGAN		Hom		pien r	s							
	(i.	(1	A) N. B) L C) I	AME/ OCAT DENT	KEY: ION: IFIC: INFO	21. ATIO	.83 N ME	THOD	core	10	ijne MVTR					
	(x.	i) SI	EQUE	NCE.	DESC	RIPT	ION:	SEQ	ID 1	NO:	42:					
AAT	ACCT:	TAG A	ACCC'	TCAG'	TC A	et P	CA G' ro Va 20	TG Co	CT Go	CT C' la L	eu C	GC C' ys Le	rg Ci	rc ro	GG GCC	53
CTG Leu -10	GCA Ala	ATG Met	GTG Val	ACC Thr	CGG Arg -5	CCT Pro	GCC Ala	TCA Ser	GCG Ala	GCC Ala 1	CCC Pro	ATG Met	GGC Gly	GGC Gly 5	CCA Pro	101
GAA Glu	CTG Leu	GCA Ala	CAG Gln 10	CAT His	GAG Glu	GAG Glu	CTG Leu	ACC Thr 15	CTG Leu	CTC Leu	TTC Phe	CAT His	GGG Gly 20	ACC Thr	CTG Leu	149
CAG Gln	CTG Leu	GGC Gly 25	CAG Gln	GCC Ala	CTC Leu	AAC Asn	GGT Gly 30	GTG Val	TAC Tyr	AGG Arg	ACC Thr	ACG Thr 35	GAG Glu	GGA Gly	CGG Arg	197
CTG Leu	ACA Thr 40	AAG Lys	GCC Ala	AGG Arg	AAC Asn	AGC Ser 45	CTG Leu	GGT Gly	CTC Leu	TAT Tyr	GGC Gly 50	CGC Arg	ACA Thr	ATA Ile	GAA Glu	245
CTC Leu 55	CTG Leu	GGG Gly	CAG Gln	GAG Glu	GTC Val 60	AGC Ser	CGG Arg	GGC Gly	CGG Arg	GAT Asp 65	GCA Ala	GCC Ala	CAG Gln	GGC Gly		290
(2)	INFO	ORMAT	NOI	FOR	SEQ	ID 1	10: 4	13:								
	(i)	(E	A) L: 3) T: 3) S:	engti (PE : [Rani		59 ba LEIC ESS:	ACII	oairs O	5							
	(ii	L) MC														

WO 99/06439 PCT/	B98/0
. 34	
(F) TISSUE TYPE: Lung (cells)	
<pre>(ix) FEATURE: (A) NAME/KEY: sig_peptide (B) LOCATION: 128250 (C) IDENTIFICATION METHOD: Von Heijne matrix (D) OTHER INFORMATION: score 9.4</pre>	
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 43:	
ATAAGGAATA AGCAATGTTT TAAAACCACT CAAAATGTTC AACTGTTGCC AAAAGATGCT	60
ACGGAGCTCT CCCTAAGCCC GTGGCCCACC CTAAATGTAA CAGGCCCATG TTTACAAACC	120
CAAATCC ATG CAT CTC AGA GGC TCC CAC ACA TAT CCT AGC TGT CCC TCC Met His Leu Arg Gly Ser His Thr Tyr Pro Ser Cys Pro Ser -40 -35 -30	169
TCA GAA CTC CGT TTG GAC AGT CTC TGG CAG CAT CAC CGG CAG CTG CTG Ser Glu Leu Arg Leu Asp Ser Leu Trp Gln His His Arg Gln Leu Leu -25 -15	217
CCT CTC TGG GTG TTC CTG CCA CTC AGC CTG GGC CCC CCT GGG Pro Leu Trp Val Phe Leu Pro Leu Ser Leu Gly Pro Pro Gly -10 -5 1	259
(2) INFORMATION FOR SEQ ID NO: 44: (i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 292 base pairs (B) TYPE: NUCLEIC ACID (C) STRANDEDNESS: DOUBLE (D) TOPOLOGY: LINEAR (ii) MOLECULE TYPE: CDNA	
<pre>(vi) ORIGINAL SOURCE: (A) ORGANISM: Homo Sapiens (F) TISSUE TYPE: Lung</pre>	
(ix) FEATURE: (A) NAME/KEY: sig_peptide (B) LOCATION: 62139 (C) IDENTIFICATION METHOD: Von Heijne matrix (D) OTHER INFORMATION: score 9.2 seq FLLMTLLLGGLTG/VA	
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 44:	
AAVAGTCTGC RRTCTTCCCA GCACAGACGT TTGGACAGAG CAGGCTCCTA AGGTCTCCAG	60
A ATG CCC GTG CCA GCM TCC TGG CCC CAC CTT CCT AGT CCT TTC CTG CTG	109

ATG ACG CTA CTG CTG GGG GGA CTC ACA GGG GTA GCT GKY GAG GAA GAG 157

Met Pro Val Pro Ala Ser Trp Pro His Leu Pro Ser Pro Phe Leu Leu

-20

									-	-						
Met -10	Thr	Leu	Leu	Leu	Gly -5	Gly	Leu	Thr	Gly	Val 1	Ala	Xaa	Glu	Glu 5	Glu	
CTG Leu	CAG Gln	GTG Val	RTT Xaa 10	CAG Gln	CCT Pro	GAC Asp	AAG Lys	TCC Ser 15	ATA Ile	TCA Ser	GTT Val	GCA Ala	GCT Ala 20	GGA Gly	AAG Lys	205
WMG Xaa	GCC Ala	ACT Thr 25	CTG Leu	CAC His	TGC Cys	ACT Thr	GTG Val 30	ACW Thr	WCC Xaa	CTG Leu	ATC Ile	CMT Xaa 35	GTG Val	GGG Gly	CCC Pro	253
ATC Ile	CAG Gln 40	TGG Trp	TTM Xaa	AGA Arg	GGA Gly	GCT Ala 45	GGA Gly	CCA Pro	GGC Gly	CGG Arg	GAA Glu 50	TTA Leu	•			292
	(ii (vi	SEC (# (E (C (# (E (C (# (E (C (E	QUENC A) LE B) TY C) ST OLECU RIGIN A) OF F) TI EATUF A) NA B) LC C) II	CE CHENGTH (PE: TRANI OPOLO JLE T AME / F OCATI CHER	HARACH: 31 NUCIDEDNE DGY: CYPE: SOURCESM: CYPE: TYPE:	CTERN 19 ba LEIC LINE LINE CE: Homo PE: I	ISTICASE FACILIDOUE EAR NA DE Sarge	CS: pairs Diens side CHOD:	s :esti : Vor :ore :q LV	n Hei 9.1 /AMLI	LVF					·
AACI	ACCC	CG G	STAGO	CTGP	C AG	CAGO	GAGCI	CGA	\GAGI				CAG Gln			55
GTT Val	TGC Cys	GTG Val	CTG Leu -15	GCC Ala	CTG Leu	GTG Val	GCT Ala	ATG Met -10	CTG Leu	CTC Leu	CTA Leu	GTT Val	TTC Phe -5	CCT Pro	ACC Thr	103
GTC Val	TCC Ser	AGA Arg 1	TCG Ser	ATG Met	GGC Gly	CCG Pro 5	AGG Arg	AGC Ser	GGG Gly	GAG Glu	CAT His 10	CAA Gln	AGG Arg	GCG Ala	TCG Ser	151
CGA Arg 15	ATC Ile	CCT Pro	TCT Ser	CAG Gln	TTC Phe 20	AGC Ser	AAA Lys	GAG Glu	GAA Glu	CGC Arg 25	GTC Val	GCG Ala	ATG Met	AAA Lys	GAG Glu 30	199
GCG Ala	CTG Leu	AAA Lys	GGT Gly	GCC Ala 35	ATC Ile	CAG Gln	ATT Ile	CCA Pro	ACA Thr 40	GTG Val	ACT Thr	TTT Phe	AGC Ser	TCT Ser 45	GAG- Glu	247

					٦(,			
TCC Ser	Asn								295
 TTC Phe		 	 						319

(2) INFORMATION FOR SEQ ID NO: 46:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 307 base pairs
 - (B) TYPE: NUCLEIC ACID
 - (C) STRANDEDNESS: DOUBLE
 - (D) . TOPOLOGY: LINEAR
- (ii) MOLECULE TYPE: CDNA
- (vi) ORIGINAL SOURCE:
 - (A) ORGANISM: Homo Sapiens
 - (F) TISSUE TYPE: Lung (cells)
- (ix) FEATURE:
 - (A) NAME/KEY: sig_peptide
 - (B) LOCATION: 251..298
 - (C) IDENTIFICATION METHOD: Von Heijne matrix
 - (D) OTHER INFORMATION: score 9

seq LISFLLLLLLLP/AR

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 46:

(2) INFORMATION FOR SEQ ID NO: 47:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 395 base pairs
 - (B) TYPE: NUCLEIC ACID
 - (C) STRANDEDNESS: DOUBLE
 - (D) TOPOLOGY: LINEAR
- (ii) MOLECULE TYPE: CDNA

(ix) FEATURE:

(A) NAME/KEY: sig_peptide

	(vi	(2	RIGIN A) OF F) TI	RGAN	SM:	Homo			5							
	(ix	() ()	EATURA) NA B) LO C) II	AME/F DCATI DENTI	ON:	132. TION	.38: ME:	3 CHOD: : sc	core	8.9	-					
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 47:																
AGAGTGCCTC AAGGGCAGAA TGGAATGAGC AGAACCCTTT GGTTTCCCGG GAAGAGCTCT													60			
CARCCTTCAC TCACCAACAC TCATTTTTTCCC TCCCACCACCACCACCACCACCACCACCACCACCACCACC													120			
TCAC	GTCC	CAG A	A ATO	G GCT	T ACA	A ACA	GT(Vai -8(l Pro	GA1 Asp	GGT Gly	TGC Cys	CGC Arg	g Ası	r GGC n Gly	CTG Leu	170
AAA Lys	TCC Ser -70	AAG Lys	TAC Tyr	TAC Tyr	AGA Arg	CTT Leu -65	TGT Cys	GAT Asp	AAG Lys	GCT Ala	GAA Glu -60	GCT Ala	TGG Trp	GGC	ATC Ile	213
GTC Val -55	CTA Leu	GAA Glu	ACG Thr	GTG Val	GCC Ala -50	ACA Thr	GCC Ala	GGG Gly	GTT Val	GTG Val -45	ACC Thr	TCG Ser	GTG Val	GCC Ala	TTC Phe -40	266
ATG Met	CTC Leu	ACT Thr	CTC Leu	CCG Pro -35	ATC Ile	CTC Leu	GTC Val	TGC Cys	AAG Lys -30	GTG Val	CAG Gln	GAC Asp	TCC Ser	AAC Asn -25	AGG Arg	314
CGA Arg	AAA Lys	ATG Met	CTG Leu -20	CCT Pro	ACT Thr	CAG Gln	TTT Phe	CTC Leu -15	TTC Phe	CTC Leu	CTG Leu	GGT Gly	GTG Val -10	TTG Leu	GGC Gly	362
	TTT Phe															395
(2)	INFO	RMA?	rion	FOR	SEQ	ID N	١٥: ٥	18:								
	(i)	() ()	QUENCA) LEB) TY	ENGTI PE: PRANI	i: 29 NUCI DEDNI	94 ba LEIC ESS:	ACII	pair: D	5							
	(ii	L) M	OLEC	ULE 1	TYPE:	: CDI	AR									
	(vi	()	RIGII A) OI F) TI	RGAN	ISM:	Homo				ine						

- (B) LOCATION: 82..138 (C) IDENTIFICATION METHOD: Von Heijne matrix
- (D) OTHER INFORMATION: score 8.9

seq WLFLVIFIKGVQC/QE

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 48:

AGC	CTG	GGA	GAAG	AGCC	CC AC	CCCC	CAGC	AT?	CCC	AGGA	GAT	CCA	TC (GGTGA	ATCAGO	60
GCT	SAAC	ACA	GAGG!	ACTC!	AC C									TTC Phe		111
			ATA Ile													159
			GTG Val													207
			TTT Phe													255
			AAG Lys													294

- (2) INFORMATION FOR SEQ ID NO: 49:
 - (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 177 base pairs
 - (B) TYPE: NUCLEIC ACID
 - (C) STRANDEDNESS: DOUBLE
 - (D) TOPOLOGY: LINEAR
 - (ii) MOLECULE TYPE: CDNA
 - (vi) ORIGINAL SOURCE:
 - (A) ORGANISM: Homo Sapiens
 - (F) TISSUE TYPE: Lung (cells)
 - (ix) FEATURE:
 - (A) NAME/KEY: sig_peptide
 - (B) LOCATION: 115..168
 - (C) IDENTIFICATION METHOD: Von Heijne matrix
 - (D) OTHER INFORMATION: score 8.6

seq VLLHVAFLPGRFG/RP

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 49:

TTCTTAAGAT TCCATGAGTT GACTAGTTCA ATTCTGTTCC ATGGGGTGTT GGCTGGTCCC TGTGTTCAGT TGGTGCCGGG GCTGGGCTGG AAGGTCCAAG ACATCTTTTC TCAC ATG 117 Met

TCT Ser	GGT Gly	ACC Thr -15	TCA Ser	GTG Val	CTT Leu	CTC Leu	CAC His -10	GTG Val	GCC Ala	TTT Phe	CTA Leu	CCT Pro -5	GGC Gly	AGG Arg	TTT Phe	165
	CGC Arg 1															177
(2)	INFO	ORMAT	rion	FOR	SEQ	ID 1	10: 5	50:							•	
	(i)	(2 (E	A) LI B) T: C) S:	engti Pe: Prani	HARAC H: 25 NUCI DEDNI DGY:	ol ba LEIC ESS:	ase p ACII DOUI	pairs O	5							
	(ii	L) MO	DLEC	JLE :	TYPE:	CDi	NA.									
	(vi	(2	A) OI	RGAN	SOURC ISM: E TYI	Homo										
	(i)	(I (C	3) L(C) II	AME/I OCAT: DENT:	KEY: ION: IFIC! INFO	 101T	.143 N MET	THOD:	core	8.6		matı /AVS/				
	(x:	L) SE	EQUE	NCE I	DESC	RIPT	ON:	SEQ	ID 8	10: 5	50:					
AATT	rctgi	CT (CACTO	GGAG!	AG G!	AGGC!	AGGG <i>I</i>	A CAG	GACCO	CAGC	AGC	ACCCA	ACC 1	rgago	GAGAA	60
GAG	CAGA	CAC (CGTG	CTCC	rg G	AATC	ACCC	A GC						CTG Leu		113
GTC Val -10	AGT Ser	CTC Leu	CTC Leu	CTC Leu	TCT Ser -5	GTT Val	GCA Ala	GTA Val	AGT Ser	Ala	Ile	AAA Lys	Glu	Leu	CCT Pro	161
GGG Gly	GTG Val	AAG Lys	AAG Lys 10	TAT Tyr	GAA Glu	GTG Val	GTT Val	TAT Tyr 15	CCT Pro	ATA Ile	AGA Arg	CTT Leu	CAT His 20	CCA Pro	CTG Leu	209
CAT His	AAA Lys	AGA Arg 25	GAG Glu	GCC Ala	AAA Lys	GAG Glu	CCA Pro 30	GAG Glu	CAA Gln	CAG Gln	GAA Glu	CGA Arg 35	CGG Arg			251
(2)	INFO	ORMA?	IION	FOR	SEQ	ID I	NO: !	51:								
	(i)	(2 (1 (4	A) LI B) T' C) S'	engt: Ype: Trani	HARAC H: 2- NUC DEDNI DGY:	49 ba LEIC ESS:	ase p ACII DOU	pair. D	5							

(ii)	MOLECULE TYPE: CDNA	
(vi)	ORIGINAL SOURCE: (A) ORGANISM: Homo Sapiens (F) TISSUE TYPE: Colon	
(ix)	FEATURE: (A) NAME/KEY: sig_peptide (B) LOCATION: 160204 (C) IDENTIFICATION METHOD: Von Heijne matrix (D) OTHER INFORMATION: score 8.6 seq ICHVSLLLQLCSS/CK	
(xi)	SEQUENCE DESCRIPTION: SEQ ID NO: 51:	
AATCCATTG	T TTTCCCTAGC AGTGTGGGGC TCTCTGACAT GCTTGTTGGT CTTGGAGGGG	60
TTCTGAGAG	G CTCTCCCTAG TCTTAAGTGC TTCTGTCTTA CCAGACCTTC CTTCTGTCCC	120
CTCACTTAAC	C AGACCATCTG CTCCGGCCAC CATTCCCAC ATG CAC ATT TGT CAT Met His Ile Cys His -15	174
GTG TCT C1 Val Ser Le -10	TA CTG CTG CAG CTT TGC TCA TCT TGC AAG AAG TCC CCA CTC eu Leu Leu Gln Leu Cys Ser Ser Cys Lys Lys Ser Pro Leu -5 1 5	222
	TG CTA CAG AAA GCC CAA AGG eu Leu Gln Lys Ala Gln Arg 10 15	249
	MATION FOR SEQ ID NO: 52: SEQUENCE CHARACTERISTICS:	
(-)	(A) LENGTH: 167 base pairs (B) TYPE: NUCLEIC ACID (C) STRANDEDNESS: DOUBLE (D) TOPOLOGY: LINEAR	
(ii)	MOLECULE TYPE: CDNA	
(vi)	ORIGINAL SOURCE: (A) ORGANISM: Homo Sapiens (F) TISSUE TYPE: Liver	
(ix)	FEATURE: (A) NAME/KEY: sig_peptide (B) LOCATION: 57155 (C) IDENTIFICATION METHOD: Von Heijne matrix (D) OTHER INFORMATION: score 8.2 seq LLCSLLFSFPFLC/HP	
(xi)	SEQUENCE DESCRIPTION: SEQ ID NO: 52:	

CATCATGAAT ATCTATCCCC AGTCCTCCAC TCTTGACTCC CTCTTGGAGT GGATAG ATG

59

Met

41

ATC Ile	TTT Phe	GCT Ala -30	GAC Asp	AGG Arg	ACA Thr	CAC His	TCC Ser -25	AGC Ser	GCT Ala	TTC Phe	ACC Thr	CTG Leu -20	ATG Met	AGG Arg	TCC Ser	107
TAT Tyr	TCT Ser -15	TTG Leu	CTT Leu	TTG Leu	TGC Cys	TCA Ser -10	CTC Leu	TTG Leu	TTC Phe	TCA Ser	TTC Phe -5	CCA Pro	TTC Phe	TTA Leu	TGC Cys	155
	CCT Pro															167
(2) INFORMATION FOR SEQ ID NO: 53: (i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 236 base pairs (B) TYPE: NUCLEIC ACID (C) STRANDEDNESS: DOUBLE (D) TOPOLOGY: LINEAR (ii) MOLECULE TYPE: CDNA (vi) ORIGINAL SOURCE: (A) ORGANISM: Homo Sapiens (F) TISSUE TYPE: Pancreas (ix) FEATURE: (A) NAME/KEY: sig_peptide (B) LOCATION: 2180 (C) IDENTIFICATION METHOD: Von Heijne matrix (D) OTHER INFORMATION: score 3 seq VCVLVGSFSASLA/GT																
	(x:	L) SE	QUE	ICE I	DESC	RIPT	ON:	SEQ	ID N	io: 5	53:					
ATC	gaaa <i>i</i>	AAG (CTCT	GCAC	Me					co Se					TA CTA al Leu -10	53
GTT Val	GGT Gly	TCC Ser	TTT Phe	TCT Ser -5	GCT Ala	TCC Ser	TTA Leu	GCA Ala	GGG Gly 1	ACT Thr	TCC Ser	AAT Asn	CTC Leu 5	TCA Ser	GAG Glu	101
					TGG Trp											149
					ATG Met											197
					AAA Lys 45											236

(2) INFORMATION FOR SEQ ID NO: 54:	
(i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 234 base pairs (B) TYPE: NUCLEIC ACID (C) STRANDEDNESS: DOUBLE (D) TOPOLOGY: LINEAR	
(ii) MOLECULE TYPE: CDNA	
<pre>(vi) ORIGINAL SOURCE: (A) ORGANISM: Homo Sapiens (F) TISSUE TYPE: Lung (cells)</pre>	
<pre>(ix) FEATURE: (A) NAME/KEY: sig_peptide (B) LOCATION: 157198 (C) IDENTIFICATION METHOD: Von Heijne matrix (D) OTHER INFORMATION: score 7.7</pre>	
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 54:	
ATTAAGCAGT CACCCATTCA GGAATGAGAA TATTATACAC CTATATATTT ATTTTGTATA	60
TATTAACAAA ATTATATATG ACATTGTAGC AGGTAGGTTT GACCCAGTTG CAGGTTGGGG	120
CTGAGATGAA ATGTAAAATT GTATGTCTCA TTCTGG ATG TTT TTA GTC TCG TGC Met Phe Leu Val Ser Cys -10	174
GTT ATC TGC ACT GGG AGC TTT GCC TTT AAT AAC TCA AAC GTT CCT CTC Val Ile Cys Thr Gly Ser Phe Ala Phe Asn Asn Ser Asn Val Pro Leu -5 1 5	222
CCC AGC AGC CGG Pro Ser Ser Arg 10	234
(2) INFORMATION FOR SEQ ID NO: 55:	
(i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 465 base pairs (B) TYPE: NUCLEIC ACID (C) STRANDEDNESS: DOUBLE (D) TOPOLOGY: LINEAR	
(ii) MOLECULE TYPE: CDNA	
<pre>(vi) ORIGINAL SOURCE: (A) ORGANISM: Homo Sapiens (F) TISSUE TYPE: Lung (cells)</pre>	
<pre>(ix) FEATURE: (A) NAME/KEY: other (B) LOCATION: 34284 (C) IDENTIFICATION METHOD: blastn (D) OTHER INFORMATION: identity 99</pre>	

PCT/IB98/01233

region 11..261 id AA171572 est

1	/ i	x)	F	'EA	TI	1R	F	
ı		^				"	-	4

- (A) NAME/KEY: other
- (B) LOCATION: 268..335
- (C) IDENTIFICATION METHOD: blastn
 (D) OTHER INFORMATION: identity 91

region 246..313 id AA171572

est

(ix) FEATURE:

- (A) NAME/KEY: other
- (B) LOCATION: 354..390
- (C) IDENTIFICATION METHOD: blastn
- (D) OTHER INFORMATION: identity 97

region 334..370

id AA171572

est

(ix) FEATURE:

- (A) NAME/KEY: other
- (B) LOCATION: 394..430
- (C) IDENTIFICATION METHOD: blastn

(D) OTHER INFORMATION: identity 94

region 376..412 id AA171572

est

(ix) FEATURE:

- (A) NAME/KEY: other
- (B) LOCATION: 300..465
- (C) IDENTIFICATION METHOD: blastn

(D) OTHER INFORMATION: identity 90

region 6..171 id AA218270

est

(ix) FEATURE:

- (A) NAME/KEY: sig_peptide
- (B) LOCATION: 262..393
- (C) IDENTIFICATION METHOD: Von Heijne matrix
- (D) OTHER INFORMATION: score 7.5

seq LMIPLLLTPITA/TS

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 55:

TTCTGGAGCC GAAGAGATC GGAAGAACAA GTGATGCCAT GGACAATATT ACCAGGCAGA 60

ACCAATTCTA CGATACCCAA GTCATCAAAC AAGAAAACGA GTCAGGCTAC GAGAGGAGAC 120

CACTGGAAAT GGAGCAGCAG CAGGCCTATC GTCCAGAAAT GAAGACAGAG ATGAAGCAAG 180

GAGCACCCAC CAGCTTCCTC CCGCCTGAAG CTTCTCAACT CGAGCCAGAC AGGCAGCAAT 240

TCCAGAGTCG AAAGAGGCCT T ATG AAG AAA ACC GGG GAC GGG GGT ACT TTG
Met Lys Lys Thr Gly Asp Gly Gly Thr Leu

-40 -39

AGC Ser	ACC Thr	GAG Glu	AGG Arg	ATA Ile -30	GGA Gly	GGG Gly	GCC Ala	GCT Ala	CTC Leu -25	CTC Leu	AGC Ser	CTC Leu	CTG Leu	CTG Leu -20	AAG Lys	339
AGG Arg	ATG Met	AAG Lys	ATG Met -15	ACT Thr	TTG Leu	ATG Met	ATA Ile	CCC Pro -10	TTG Leu	TTG Leu	CTA Leu	TTG Leu	ACA Thr -5	CCT Pro	ATA Ile	387
ACT Thr	GCG Ala	ACC Thr 1	TCC Ser	AMT Xaa	TCA Ser	AGG Arg 5	TGG Trp	CCC Pro	GAG Glu	ATC Ile	GGA Gly 10	GTA Val	GTG Val	GCT Ala	ATC Ile	435
	TCA Ser															465
(2)	(ii) (vi	SEC (# (# (# (# (# (# (# (# (# (# (# (# (#	QUENCAL LE	CE CE ENGTE (PE: TRANI OPOLO JLE S RGANI ISSUE RE: AME/I OCATI THER	HARACH 14 40 NUCIDEDNE DGY: TYPE: SOURCE TYPE E TYPE KEY: ION: INFO	CTERIOS backers before the control of the control o	STICES ACTIONS ACTIONS ACTIONS	CS: Dairs Dairs Oiens (cell	ils) : Vor core eq XI	7.5 [LLAC	GWCPI					
AAC	TTG	rgc (CCAG	AGCT(CC A	GCCAG	CAGTO	G AGO	CATGO	SAGC	TGG	GCGG	CCC 1	AGGG:	raa	57
ATG Met -25	GGC Gly	TTC Phe	TTT Phe	CTT Leu	CCC Pro -20	CAT His	GGC Gly	ATC Ile	TCA Ser	GAC Asp -15	GCC Ala	KGA Xaa	ATA Ile	CTC Leu	CTG Leu -10	105
	GGC Gly															153
	CCT Pro															201
	AGT Ser 25															249

Met Trp Leu Arg Pro Gly Ser

-45

TGC TGG AGT ACG AGG GAG CCA AGA AGG GCT CCA AGG ACC TCT GCC TCT Cys Trp Ser Thr Arg Glu Pro Arg Arg Ala Pro Arg Thr Ser Ala Ser

TCT CTG AGC TCG TTC TTA GGC CCC TCT GCC GTC TGC ACG CTC CTT TCC Ser Leu Ser Ser Phe Leu Gly Pro Ser Ala Val Cys Thr Leu Leu Ser

AGC CAC CCG GCC TCC CGA TGC CGG CCT AGC ACG TTC CTC GCG CCA GGC Ser His Pro Ala Ser Arg Cys Arg Pro Ser Thr Phe Leu Ala Pro Gly

1

TTT TGC ATC TGC CCT TCC CAC TGT CTT TCG TGT GCA AAC GCC ACA GAT

-35

-20

-5

46

Phe 10	Cys	Ile	Cys	Pro	Ser 15	His	Cys	Leu	Ser	Cys 20	Ala	Asn	Ala	Thr	Asp 25	
	GCG Ala															313
(2)	INFO	ORMA'	rion	FOR	SEQ	ID I	NO:	58:								
	(i)	() ()	A) LI B) T' C) S'	CE CE ENGTE YPE: TRANI DPOLO	H: 20 NUCI DEDNI	09 ba LEIC ESS:	ase p ACII	pair: D	5							
	(ii	L) MO	OLEC	ULE	YPE	: CDI	NA									
	iv)	(2	A) O	NAL S RGANI ISSUI	ISM:	Home		pien:	5							
	(i)	() ()	B) L(C) II	RE: AME/I OCATI DENTI THER	ION: [FIC	66. ATIO	.182 N ME'	THOD	: Voi core eq Ll	7.4	_					
	(xi	i) Si	EQUE	NCE I	DESCI	RIPT	ION:	SEQ	ID I	: ON	58:					
AATT	racc <i>i</i>	ATG '	rtgc	CTCA	AT TO	CTTA	GGAA	G AC'	rtct'	TTTT	AAT	TGAG	AGG 1	CTT	CTTGC	r 60
CCT	TT AT	rg To	CA G er G	AA GO	Ly Me	IG G et V 35	TC A	CA T'	TG C'	eu T	CT T' hr Pi 30	TT To	CT TO	GT T ýs Le	TA TGO eu Trp -25	p.
ACT Thr	GAT Asp	GAC Asp	TCA Ser	TTT Phe -20	ATG Met	AGT Ser	KAT Xaa	CTA Leu	AAC Asn -15	GTC Val	TTG Leu	TTC Phe	CTT Leu	CTT Leu -10	TCT Ser	158
CTC Leu	TTC Phe	TGC Cys	CGT Arg -5	CTC Leu	TAT Tyr	CAT His	GGG Gly	ACT Thr 1	ATT Ile	TTC Phe	TTT Phe	CTG Leu 5	CTA Leu	GCA Ala	CTT Leu	206
CTG Leu																209
(2)	INFO	ORMA	TION	FOR	SEQ	ID	NO:	59:								
	(i)	(A) L B) T C) S	CE CI ENGT: YPE: TRAN	H: 2	27 b LEIC ESS:	ase ACI DOU	pair D	s							

(ii) MOLECULE TYPE: CDNA

- (vi) ORIGINAL SOURCE: (A) ORGANISM: Homo Sapiens
 - (F) TISSUE TYPE: Lung
 - (ix) FEATURE:
 - (A) NAME/KEY: sig_peptide
 - (B) LOCATION: 69..110
 - (C) IDENTIFICATION METHOD: Von Heijne matrix
 - (D) OTHER INFORMATION: score 7.3

seq LILGLPLCRPLWI/QR

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 59:

AAGTCTGTCC GTGGATACTG TGAACATCAG GCTACTCGGC CGGGCTCCTG CGCTCAGGGC

TTCGAGAA ATG CTC ATT TTG GGG CTG CCC CTC TGC CGG CCT CTC TGG ATT Met Leu Ile Leu Gly Leu Pro Leu Cys Arg Pro Leu Trp Ile -10

CAG AGG GCA GCC GCT GCT CCT TTT GTT TTG TGG GCC TGG CTC TGG GCC 158 Gln Arg Ala Ala Ala Pro Phe Val Leu Trp Ala Trp Leu Trp Ala 1

CGG AGC AGC ACC TCC CTG GGG AGG CCG CCT TTC CTT CCG CGG CTT CTT 206 Arg Ser Ser Thr Ser Leu Gly Arg Pro Pro Phe Leu Pro Arg Leu Leu

CCG TCT CCT CCT GAC CCC GAG 227 Pro Ser Pro Pro Asp Pro Glu 35

(2) INFORMATION FOR SEQ ID NO: 60:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 162 base pairs
 - (B) TYPE: NUCLEIC ACID
 - (C) STRANDEDNESS: DOUBLE
 - (D) TOPOLOGY: LINEAR
- (ii) MOLECULE TYPE: CDNA
- (vi) ORIGINAL SOURCE:
 - (A) ORGANISM: Homo Sapiens
 - (F) TISSUE TYPE: Lung
- (ix) FEATURE:
 - (A) NAME/KEY: sig_peptide
 - (B) LOCATION: 16..90
 - (C) IDENTIFICATION METHOD: Von Heijne matrix
 - (D) OTHER INFORMATION: score 7.3

seq HFILLVLPCLIFS/HF

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 60:

ACATCATCTT GCTCT ATG TAT ATA TAT TTT TTT GTC TTA TGT GKK CTG TCT 51 Met Tyr Ile Tyr Phe Phe Val Leu Cys Xaa Leu Ser -25 -20

48

		ATT Ile														99
TTA Leu	TTT Phe 5	TTA Leu	TTT Phe	TAC Tyr	AGT Ser	GCC Ala 10	TTA Leu	TTA Leu	GAT Asp	ATA Ile	CCT Pro 15	CTT Leu	TTT Phe	TTC Phe	AAG Lys	147
		CTG Leu														162
(2)	INFC	RMAT	ION	FOR	SEQ	ID N	10: 6	51:								
	(i)	(E	L) LE B) TY C) ST	NGTH PE: RAND	: 31 NUCI EDNE	TERI 18 ba LEIC LSS: LINE	se p ACIO DOUB	pairs)	5							
	(ii	.) MC	LECU	LE 1	YPE	CD	JA									
	(vi		() OF	GANI	SM:	CE: Homo PE: I		oiens	5							
	(ix	(E	A) NA B) LC C) IC	ME/F CATI ENTI	ON:	sig_ 112. ATION DRMAT	.23 N ME	7 THOD: : sc	core		-					
	(xi	.) SE	QUEN	ICE E	DESC	RIPT	ON:	SEQ	ID t	NO: 6	61:					
AATA	CTTI	CT C	CTCI	cccc	CT C	rccc;	AAGC	A CA	rctg!	AGTT	GCT	GCCT	STT (CTTC	ACACTT	60
AGCT	CCAF	AAC (CATO	SAAA	AA T	rgcci	AAGT!	A TAI	AAAG(CTTC	TCA	AGAA'	rga (G GAT t Asp	117
TCT Ser -40	AGG Arg	GTG Val	TCT Ser	TCA Ser	CCT Pro -35	GAG Glu	AAG Lys	CAA Gln	GAT Asp	AAA Lys -30	GAG Glu	AAT Asn	TTC Phe	GTG Val	GGT Gly -25	165
		AAT Asn														213
		CTG Leu														261
TTG Leu	AAG Lys 10	ATC Ile	ATT Ile	AAG Lys	GMG Xaa	TAT Tyr 15	GAA Glu	CGT Arg	GMT Xaa	GTT Val	GTA Val 20	TTC Phe	CGT Arg	CTG Leu	GGA Gly	309
CGA	CAT	GGG														318

Arg His Gly 25

(2)	INFORM	ATION FOR SEQ ID NO: 62:	
		EQUENCE CHARACTERISTICS: (A) LENGTH: 154 base pairs (B) TYPE: NUCLEIC ACID (C) STRANDEDNESS: DOUBLE (D) TOPOLOGY: LINEAR	
	(ii) N	MOLECULE TYPE: CDNA	
		ORIGINAL SOURCE: (A) ORGANISM: Homo Sapiens (F) TISSUE TYPE: Thyroid	
	1	FEATURE: (A) NAME/KEY: sig_peptide (B) LOCATION: 101148 (C) IDENTIFICATION METHOD: Von Heijne matrix (D) OTHER INFORMATION: score 7 seq LFCVVLCLSPTSY/CY	
	(xi) \$	SEQUENCE DESCRIPTION: SEQ ID NO: 62:	
TGT	TTTCCT	TCTAGAGGTC CATATGTTTA CAGGCAAATT CCTACGTACA CTATATACTT	60
CCAI	ACCATCC	CTTCCTTTCC CTTTCAGAAA TAGTAGCACA ATG TGT ATA CTG TTC Met Cys Ile Leu Phe -15	115
TGT Cys	GTT GTG Val Val -10	G TTA TGT TTG TCT CCA ACA TCT TAT TGT TAT 1 Leu Cys Leu Ser Pro Thr Ser Tyr Cys Tyr -5 1	154
(2)	INFORM	ATION FOR SEQ ID NO: 63:	
		EQUENCE CHARACTERISTICS: (A) LENGTH: 163 base pairs (B) TYPE: NUCLEIC ACID (C) STRANDEDNESS: DOUBLE (D) TOPOLOGY: LINEAR	
	(ii) 1	MOLECULE TYPE: CDNA	
		ORIGINAL SOURCE: (A) ORGANISM: Homo Sapiens (F) TISSUE TYPE: Lung (cells)	
		FEATURE: (A) NAME/KEY: sig_peptide (B) LOCATION: 80136 (C) IDENTIFICATION METHOD: Von Heijne matrix (D) OTHER INFORMATION: score 7	

seq ETLLCLGSSCCQC/RI

(xi)	SEQUENCE	DESCRIPTION:	SEO	ID NO:	63:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 63:	
GGCAGCATTC AATCCAGGAA GTGCCAGCAT CACATGGTGA CTTCTGGTAG CTGTAACATT	60
TAGTGACTGT CTCCATGTC ATG CAC AGG GGT GAC ATC GAG ACC CTC TTA TGC Met His Arg Gly Asp Ile Glu Thr Leu Leu Cys -15 -10	112
CTG GGA AGT TCC TGC TGT CAA TGC AGA ATA TTC TCT TTT TTT TTT Leu Gly Ser Ser Cys Cys Gln Cys Arg Ile Phe Ser Phe Phe Phe -5 1 5	160
TTT Phe	163
(2) INFORMATION FOR SEQ ID NO: 64:	
(i) SEQUENCE CHARACTERISTICS:(A) LENGTH: 411 base pairs(B) TYPE: NUCLEIC ACID(C) STRANDEDNESS: DOUBLE(D) TOPOLOGY: LINEAR	
(ii) MOLECULE TYPE: CDNA	
<pre>(vi) ORIGINAL SOURCE: (A) ORGANISM: Homo Sapiens (F) TISSUE TYPE: Lung</pre>	
<pre>(ix) FEATURE: (A) NAME/KEY: sig_peptide (B) LOCATION: 331387 (C) IDENTIFICATION METHOD: Von Heijne matrix (D) OTHER INFORMATION: score 7</pre>	
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 64:	
AAGCCTAGGT GTGGCGCCCC GACCGGACTT TCACTTCTGG CCAGCCCTTT CCCCACCTGG	60
GCGCGGGASS GGTGCCAGTC TTTAAACAAC CTCTCGATGG GTCCCACGAA GATGTTTCCA	120
GACCCTTGGA ATGCCAAGTT CAAGTTTAGC TATGTCTCGC GGAGAGGCCG GTGGAAGAAG	180
CAACGAGAAT GAAGCACCCC AGTTCTCTGC TGAGCACATG GGCATCTGCA ATAAAGATTT	240
ARTITICCCAG CITCICCIGA AGCICGGIAI GGCCACAACA CIAAATICIG CCCGAGGAGA	300
TWGAGCAAAA TAGTATGGGA CTTCCAAGAA ATG TTT TTA AAG TCA GGG GCA GGC Met Phe Leu Lys Ser Gly Ala Gly	354

CTT TCT TCA TGC CTT CTT CCT CTT TGC TGG CTG GAA CGC AAA GAC CAT Leu Ser Ser Cys Leu Leu Pro Leu Cys Trp Leu Glu Arg Lys Asp His -10 5

WO 99/06439 F	PCT/IB98/012
51	
GGC AGG AGG Gly Arg Arg	411
(2) INFORMATION FOR SEQ ID NO: 65:	
 (i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 368 base pairs (B) TYPE: NUCLEIC ACID (C) STRANDEDNESS: DOUBLE (D) TOPOLOGY: LINEAR 	
(ii) MOLECULE TYPE: CDNA	
<pre>(vi) ORIGINAL SOURCE: (A) ORGANISM: Homo Sapiens (F) TISSUE TYPE: Lung (cells)</pre>	
<pre>(ix) FEATURE: (A) NAME/KEY: sig_peptide (B) LOCATION: 9122 (C) IDENTIFICATION METHOD: Von Heijne matrix (D) OTHER INFORMATION: score 7</pre>	
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 65:	
ACCTCCTA ATG GCA AAT GCA ATA ATT AAG AAG CCT TGT GCA ATG CCG Met Ala Asn Ala Ile Ile Lys Lys Pro Cys Ala Met Pro -35 -30	
CAG CCT CAC ACA GGG AAT CTC CTG TGG CCA CCT CTT GTC ATG GTG TG Gln Pro His Thr Gly Asn Leu Leu Trp Pro Pro Leu Val Met Val T-20 -15 -10	

CTC GGG CTG TTG CCT CTT TTC TCT GGG CCC CAC TTG CAA GCC GTT CAG Leu Gly Leu Leu Pro Leu Phe Ser Gly Pro His Leu Gln Ala Val Gln

CAT CTG GCC CTT GCC TAC CTC CCT CTG AAC TCT GTG GTT CTT GCA CAT

His Leu Ala Leu Ala Tyr Leu Pro Leu Asn Ser Val Val Leu Ala His

AAT AGT CCA GCC ATA CTG AAT GTA TGG TTA ACT CTG AGA TGC CCT CTC

Asn Ser Pro Ala Ile Leu Asn Val Trp Leu Thr Leu Arg Cys Pro Leu

CCT TAC AGG ATC TGC CTC TGG CCC TTT GAA CAT GCT TTT CCG TCT ATT

Pro Tyr Arg Ile Cys Leu Trp Pro Phe Glu His Ala Phe Pro Ser Ile

AGA AAC ACA CAC TCC TGC CTG TCT TCC TCC TGC TGC TGC CCT GCT TCA Arg Asn Thr His Ser Cys Leu Ser Ser Ser Cys Cys Cys Pro Ala Ser

50

30

GCC CCT CTT TTA GTT GAT TAC CTC TTC TTT

Ala Pro Leu Leu Val Asp Tyr Leu Phe Phe

1

194

242

290

368

(2)	INFO	RMAT	CION	FOR	SEQ	ID N	10: 6	56:								
	(i)	(E	A) LE 3) TY C) S1	ENGTH (PE: TRAND	I: 40 NUCI EDNE	CTERI 07 ba LEIC ESS: LINE	se p ACII DOUB	oairs O	5							
	(ii	.) MC	DLEC	JLE I	YPE	CD1	IA									
	(v)		A) OF	RGANI	SM:	CE: Homo PE: I	-	oiens	5							
		(E	A) NA B) LO C) II D) OT	AME/F OCATI DENTI THER	ON: FICA INFO	ORMAT	74 N MET	HOD: se	: Vor core eq L(6.9 PLLI	LLLPI					
AACO	CCAGO	ятс с	CCA					Pro I	CCG (Pro I -15				Pro 1			50
		CTG Leu														98
		CCT Pro														146
		GGA Gly														194
		GGG Gly														242
GTG Val	CAG Gln	TAT Tyr	TTT Phe 60	GGG	GAA Glu	ATT Ile	GGG Gly	CTG Leu 65	GGA Gly	ACG Thr	CCT Pro	CCA Pro	CAA Gln 70	AAC Asn	TTC Phe	290
		GCC Ala 75				Gly										338
		CAC His														386
	Lys	CCT Pro				Ser										407

(2) INFORMATION FOR SEQ ID NO: 67:	
(i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 395 base pairs (B) TYPE: NUCLEIC ACID (C) STRANDEDNESS: DOUBLE (D) TOPOLOGY: LINEAR	
(ii) MOLECULE TYPE: CDNA	
<pre>(vi) ORIGINAL SOURCE: (A) ORGANISM: Homo Sapiens (D) DEVELOPMENTAL STAGE: Fetal (F) TISSUE TYPE: liver</pre>	
<pre>(ix) FEATURE: (A) NAME/KEY: sig_peptide (B) LOCATION: 264389 (C) IDENTIFICATION METHOD: Von Heijne matrix (D) OTHER INFORMATION: score 6.5</pre>	
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 67:	
ATATAGTTTT GAAAGACTAC TTGAAAAACA TAATCAAATG ATACTTGGAG TTTGTTCCGA	60
AAGGGTATTC TTTCCTCTCT CTTCCCACAA CAGTGTGCGC TTCCTNNNNA TAGTTGATAT	120
ACCTGTGCTT TAAAATAGCA CAGCTACTAG TAGAGGCCAT TTTATACCAA AATCATTTCC	180
CTTCATGTTC TGTGGTACAT CAGTTTGGCA GTAGAGGTTA CAGAGTTTGA AATCAAAAGG	240
AGCATTGGTT CCTTCAGGGA AAA ATG ATA CCA ATC TAC CAA AAT AAA AGC CAA Met Ile Pro Ile Tyr Gln Asn Lys Ser Gln -40 -35	293
ACA GAC TCT CAT TGT TCT TTA TCC CAC AAG GGG CTT GCC TTT TTG AAG Thr Asp Ser His Cys Ser Leu Ser His Lys Gly Leu Ala Phe Leu Lys -30 -25 -20	341
GTG TGG TTA ATT TTG ATA GGA CTC TTC TCT CTA ACA GGG TTA GTG GCT Val Trp Leu Ile Leu Ile Gly Leu Phe Ser Leu Thr Gly Leu Val Ala -15 -5	389
GGG AAT Gly Asn 1	395
(2) INFORMATION FOR SEQ ID NO: 68:	
(i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 205 base pairs (B) TYPE: NUCLEIC ACID (C) STRANDEDNESS: DOUBLE (D) TOPOLOGY: LINEAR	

107

PCT/IB98/01233

WO 99/004	139											PC1/IB	70/01
٠						5	4						
(ii)	MOLECU	LE TYPE	: CD	NΑ									
(vi)		AL SOUR GANISM: SSUE TY	Homo						•				
(ix)	(B) LO	E: ME/KEY: CATION: ENTIFIC HER INF	.86 ATION	.145 N MET	THOD	: Vor core eq Ll	6.5						
(xi)	SEQUEN	CE DESC	RIPT	ON:	SEQ	ID 1	10: (68:					
AGGTGTCTG	C CCTGC	TTATG G	GATTO	STGC!	A GT	rgaco	CTGA	GCAG	GAGG!	ATG (GAGC!	CACTC	60
CTCTTTAAA	C CAAGG	GCCCT G	AGCC			CTT Leu							112
TCA GGC T Ser Gly S -10	CA ACT (er Thr (TGC CCT Cys Pro	GGA Gly -5	CCC Pro	TGT Cys	TCC Ser	TGT Cys	GGA Gly 1	TCC Ser	TTG Leu	AGG Arg	AGC Ser 5	160
CCT CCT G	GG CCT (GTG ACT Val Thr 10	GAT Asp	AAA Lys	CCC Pro	CTC Leu 15	CCC Pro	CTG Leu	CCG Pro	CCC Pro	CAG Gln 20		205
(ii) (vi) (ix)	SEQUENCI (A) LEI (B) TYI (C) STI (D) TOI MOLECUI ORIGINA (A) ORI (F) TI: FEATUR (A) NAI (B) LOI (C) IDI	E CHARANGTH: 2 PE: NUC RANDEDN POLOGY: LE TYPE AL SOUR GANISM: SSUE TY E: ME/KEY: CATION: ENTIFIC HER INF	CTER: 45 ba LEIC ESS: LINI : CDM CE: Homo PE: (60. ATION ORMA	ISTICASE IN ACII DOUBLEAR NA DO Sag Color 182 N ME' TION	CS: pair: D BLE pien n tide THOD: s	s : Vo: core eq I	6.4 ALIP	LFST					
(**)	SEQUENT	CE DESC	NIF1.	TON:	೨೬Q	ID	NO:	:עס					

AATTCTAGGC TGGTCACTAC TCCGAGCCTG TDWCGTTTGC GGCAGCCAGG CCGTCGACG

ATG CCC AGT GAA ACT CTC TGG GAA ATT GCA AAA GCT GAA GTG GAA AAA Met Pro Ser Glu Thr Leu Trp Glu Ile Ala Lys Ala Glu Val Glu Lys

	,,,,,,								_	_					. 01/12/	0.011
	•								3	5						
	-40					-35					-30					
AGG Arg -25	GGA Gly	ATT Ile	AAT Asn	GGA Gly	ART Xaa -20	YAA Xaa	GGT Gly	GAT Asp	Gly	GCT Ala -15	GAA Glu	ATT Ile	GCA Ala	TTA Leu	ATT Ile -10	155
CCC Pro	CTT Leu	TTT Phe	TCC Ser	ACT Thr -5	WCA Xaa	GCT Ala	TTT Phe	GCA Ala	ATT Ile 1	DCC Xaa	CAG Gln	ATA Ile	GTK Val 5	TCA Ser	CTG Leu	203
GGC Gly	ATC Ile	GTC Val 10	GAC Asp	GGC Gly	AGT Ser	DCT Xaa	CCA Pro 15	CCA Pro	RGA Xaa	TCC Ser	AGG Arg	ACC Thr 20	CCG Pro			245
(2)	INFO				_											
	(i)	(I (I	A) LE B) T' C) ST	ENGTI (PE: (RANI	IARAC I: 23 NUCI DEDNI DGY:	84 ba LEIC ESS:	ase p ACII DOUE	pairs D	6							
	(ii	i) MO	DLEC	JLE 1	YPE	CDI	IA									
	(vi	(2	A) OI	RGAN	SOUR(SM: E TY!	Homo										
	(1)	(I	A) NA B) L(C) I(AME/E DCATE DENTE	(EY: ION: IFIC! INF(61 TION	171 1 ME1	THOD:		6.4	-					
	(x)	i) Si	EQUE	NCE I	DESCI	RIPT	ON:	SEQ	ID N	io: 7	70:					
ATG	CTAC	CCA (GGAC/	AATG	CT AC	GGGG	CACA	r GG	GCAC	AGCA	TGA	ATA	ACT (GCAT	CTCAGA	60
ATG Met	GAA Glu	TGG Trp -35	TTA Leu	CGC Arg	CCC Pro	TCC Ser	CAA Gln -30	ATC Ile	TCC Ser	TTC Phe	TAT Tyr	CCG Pro -25	GGT Gly	TAC Tyr	AGC Ser	108
AAG Lys	GAA Glu -20	AGG Arg	CTC Leu	CGT Arg	TTG Leu	GTG Val -15	CTA Leu	CTA Leu	TGC Cys	ATG Met	TCC Ser -10	CTA Leu	ACC Thr	TTT Phe	CTA Leu	156
GCA Ala -5	CTT Leu	TCT Ser	ACT Thr	CTC Leu	CGC Arg 1	TTT Phe	TTA Leu	ACA Thr	CAG Gln 5	AGA Arg	GTG Val	CAG Gln	ATG Met	CAG Gln 10	GCT Ala	204

234

(2) INFORMATION FOR SEQ ID NO: 71:

GGG TGC CCT CTG CGG AGT CCA CGC CTC TGG

Gly Cys Pro Leu Arg Ser Pro Arg Leu Trp

(A) LE (B) TY (C) ST	E CHARACTERI NGTH: 163 ba PE: NUCLEIC RANDEDNESS: PPOLOGY: LINE	se pairs ACID DOUBLE					
(ii) MOLECU	LE TYPE: CDN	Α					
	NAL SOURCE: RGANISM: Homo RSSUE TYPE: L	•					
(B) LC (C) ID	RE: MME/KEY: sig_ DCATION: 41 DENTIFICATION THER INFORMAT	94 METHOD: ION: sco	ore 6.3	.jne matı			
(xi) SEQUEN	CE DESCRIPTI	ON: SEQ	ID NO: 7	1:			
AATCATTTCC TCCTC	CAGATT ACCAAG	CAAG AAC	AGCTAAA	ATG AAA Met Lys		lle	55
CAT CTT ACT CTT His Leu Thr Leu -10	CTT GCT CTC Leu Ala Leu	CTT TCT (Leu Ser '	GTA AAC Val Asn	ACA GGT Thr Gly	AAG GAA Lys Glu 1	TAT Tyr	103
TTT TAC ATT TTA Phe Tyr Ile Leu 5							151
TCA GCC GGC CAG Ser Ala Gly Gln 20							163
(2) INFORMATION	FOR SEQ ID N	10: 72:					
(A) LE (B) TY (C) SI	CE CHARACTERI ENGTH: 155 ba PE: NUCLEIC FRANDEDNESS: PPOLOGY: LINE	se pairs ACID DOUBLE				·	
(ii) MOLECU	JLE TYPE: CDN	IA					
(A) OF	NAL SOURCE: RGANISM: Homo ISSUE TYPE: I						
(B) LC (C) II	RE: AME/KEY: sig_ OCATION: 21 DENTIFICATION THER INFORMAT	.107 N METHOD: TION: sc	ore 6.3	ijne mat			

(xi) 3	SEQUENCE	DESCRIPTION:	SEO	ID NO:	72:
--------	----------	--------------	-----	--------	-----

ACACTACTAT TTTAAGGAAA ATG GAT GTG TCA GCC AGC AAG CCA GTG GCA GAG Met Asp Val Ser Ala Ser Lys Pro Val Ala Glu -25 -20	53
TCT TGG TCT CCA GGC TCC CTG CCT CTT GCA CTG ACT CTT TCT CTT TCA Ser Trp Ser Pro Gly Ser Leu Pro Leu Ala Leu Thr Leu Ser Leu Ser -15 -5	101
ACC TCC CTG CAT GAC AGC TGG AAA GAG CCC ATC CCT AAT CTT CAC CAA Thr Ser Leu His Asp Ser Trp Lys Glu Pro Ile Pro Asn Leu His Gln 1 5 10	149
CCG GCG Pro Ala 15	155
(2) INFORMATION FOR SEQ ID NO: 73: (i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 149 base pairs (B) TYPE: NUCLEIC ACID (C) STRANDEDNESS: DOUBLE (D) TOPOLOGY: LINEAR (ii) MOLECULE TYPE: CDNA (vi) ORIGINAL SOURCE: (A) ORGANISM: Homo Sapiens (F) TISSUE TYPE: Liver (ix) FEATURE: (A) NAME/KEY: sig_peptide (B) LOCATION: 33116 (C) IDENTIFICATION METHOD: Von Heijne matrix (D) OTHER INFORMATION: SCORE 6.2 seq IQTALLGLPXAWA/SS (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 73:	
ACCTGGTTGT AACCCAGCCT TCTTCCTTAC AG ATG GGG GTG AGG GTT GGG GTA	53
TCC CTT AGA GCC TGG TGT GTC TTC ATA CAA ACA GCT TTG CTG GGG CTG Ser Leu Arg Ala Trp Cys Val Phe Ile Gln Thr Ala Leu Leu Gly Leu -20 -15 -10	101
CCT ARA GCC TGG GCC AGC AGT GGG GTA GTA TCA AGT ACA GGG CCG GGG Pro Xaa Ala Trp Ala Ser Ser Gly Val Val Ser Ser Thr Gly Pro Gly -5 1 5 10	149

(2) INFORMATION FOR SEQ ID NO: 74:

(i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 106 base pairs (B) TYPE: NUCLEIC ACID (C) STRANDEDNESS: DOUBLE (D) TOPOLOGY: LINEAR	
(ii) MOLECULE TYPE: CDNA	
<pre>(vi) ORIGINAL SOURCE: (A) ORGANISM: Homo Sapiens (F) TISSUE TYPE: Lung (cells)</pre>	
<pre>(ix) FEATURE: (A) NAME/KEY: sig_peptide (B) LOCATION: 2085 (C) IDENTIFICATION METHOD: Von Heijne matrix (D) OTHER INFORMATION: score 6.1</pre>	
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 74:	
CCATGAAACC ATCAAGATT ATG ATT ATA TCC ATC ATC CCT AGA AGT TTC TTC Met Ile Ile Ser Ile Ile Pro Arg Ser Phe Phe -20 -15	52
CTA CTG CTT TGT ATT CCC TTT CTT ACC CTC CTC TTG TAT ACA TAC CCC Leu Leu Leu Cys Ile Pro Phe Leu Thr Leu Leu Leu Tyr Thr Tyr Pro -10 -5 1 5	100
CCC AGG Pro Arg	106
(2) INFORMATION FOR SEQ ID NO: 75:	
(i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 433 base pairs (B) TYPE: NUCLEIC ACID (C) STRANDEDNESS: DOUBLE (D) TOPOLOGY: LINEAR	
(ii) MOLECULE TYPE: CDNA	
<pre>(vi) ORIGINAL SOURCE: (A) ORGANISM: Homo Sapiens (F) TISSUE TYPE: Thyroid</pre>	
<pre>(ix) FEATURE: (A) NAME/KEY: sig_peptide (B) LOCATION: 272358 (C) IDENTIFICATION METHOD: Von Heijne matrix (D) OTHER INFORMATION: score 5.9</pre>	
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 75:	

. 39	
ATGGNMAAGA TGCTCGCAGG CTGCTTTCTG CTGATCCTCG GACAGATCGT CCTCCTCCCT	120
GCCGAGGCCA GGGAGCGGTC ACGTGGGAGG TCCATCTCTA GGGGCAGACA CGCTCGGACC	180
CACCCGCAGA CGGCCCTTCT GGAGAGTTCC TGTGAGAACA AGCGGGCAGA CCTGGTTTTC	240
ATCATTGACA GCTCTCGCAG TGNNAACACC C ATG ACT ATG CAA AGG TCA AGG Met Thr Met Gln Arg Ser Arg -25	292
AGT TCA TCG TGG ACA TCT TGC AAT TCT TGG ACA TTG GTC CTG ATG TCA Ser Ser Trp Thr Ser Cys Asn Ser Trp Thr Leu Val Leu Met Ser -20 -15 -10	340
CCC GAG TGG GCC YTG CTC CAA TAT GGC AGC ACT GTC AAG AAT GAG TTC Pro Glu Trp Ala Leu Leu Gln Tyr Gly Ser Thr Val Lys Asn Glu Phe -5 1 5 10	388
TCC YSM AAG ACC TTC AAG AGG AAG TCC GAG GTG GAG CGT GCT GTC Ser Xaa Lys Thr Phe Lys Arg Lys Ser Glu Val Glu Arg Ala Val 15 20 25	433
(2) INFORMATION FOR SEQ ID NO: 76: (i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 328 base pairs (B) TYPE: NUCLEIC ACID (C) STRANDEDNESS: DOUBLE (D) TOPOLOGY: LINEAR (ii) MOLECULE TYPE: CDNA (vi) ORIGINAL SOURCE: (A) ORGANISM: Homo Sapiens (F) TISSUE TYPE: Lung (cells)	
(ix) FEATURE:	

- FEATURE:

 (A) NAME/KEY: sig_peptide

 (B) LOCATION: 239..307

 (C) IDENTIFICATION METHOD: Von Heijne matrix
- (D) OTHER INFORMATION: score 5.8

seq LCSLMASISPTLT/AV

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 76:

CCTTTCTGGC TGTGGCTGAC TCCTCGGACA CTCTTCCTTC TCATCTTCCT TCGGCTTCTA	60
GACCACAGTC TTCTTCTTTG CTTTTTGTCC TTCCAGGCAG ATTGGCAATG AAAGTTGTGC	120
CGCTCGATGT CCCTGAACTG CCCCAGGCCT CCTCCTCCCC TCAAGGCACT CCACGGCTAG	180
CCACCTGGCA GCCTGGGCTG CTCACTAGAA GGCTACTGCC CTCTACCCCA CTGCTTTC	238
ATG ATC ACC CTC CCT CAG ACC TCC AGC CTG CTC TGT AGC CTC ATG GCC Met Ile Thr Leu Pro Gln Thr Ser Ser Leu Leu Cys Ser Leu Met Ala -20 -15 -10	286

- 5 1 5	328							
(2) INFORMATION FOR SEQ ID NO: 77:								
(i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 334 base pairs (B) TYPE: NUCLEIC ACID (C) STRANDEDNESS: DOUBLE (D) TOPOLOGY: LINEAR								
(ii) MOLECULE TYPE: CDNA-								
<pre>(vi) ORIGINAL SOURCE: (A) ORGANISM: Homo Sapiens (F) TISSUE TYPE: Liver</pre>								
<pre>(ix) FEATURE: (A) NAME/KEY: sig_peptide (B) LOCATION: 101289 (C) IDENTIFICATION METHOD: Von Heijne matrix (D) OTHER INFORMATION: score 5.7</pre>								
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 77:								
AAGTTAATCA GTCACTGCCC CTGGTCTGGG CAAGGGGCAG TTGGTGAACT TGCTGCCTCC								
AAGTTAATCA GTCACTGCCC CTGGTCTGGG CAAGGGGCAG TTGGTGAACT TGCTGCCTC	C 60							
AAGTTAATCA GTCACTGCCC CTGGTCTGGG CAAGGGGCAG TTGGTGAACT TGCTGCCTC AGAGAACCTT CCCTGGTGTG GAGGCAGCCA GGGACCCAGG ATG CTC CGG ACC TGT Met Leu Arg Thr Cys -60	C 60							
AGAGAACCTT CCCTGGTGTG GAGGCAGCCA GGGACCCAGG ATG CTC CGG ACC TGT Met Leu Arg Thr Cys								
AGAGAACCTT CCCTGGTGTG GAGGCAGCCA GGGACCCAGG ATG CTC CGG ACC TGT Met Leu Arg Thr Cys -60 TAC GTG CTC TGT TCC CAA GCT GGT CCC CCC TCC AGG GGC TGG CAG TCC Tyr Val Leu Cys Ser Gln Ala Gly Pro Pro Ser Arg Gly Trp Gln Ser	115							
AGAGAACCTT CCCTGGTGTG GAGGCAGCCA GGGACCCAGG ATG CTC CGG ACC TGT Met Leu Arg Thr Cys -60 TAC GTG CTC TGT TCC CAA GCT GGT CCC CCC TCC AGG GGC TGG CAG TCC Tyr Val Leu Cys Ser Gln Ala Gly Pro Pro Ser Arg Gly Trp Gln Ser -55 -50 -45 CTG AGC TTT GAT GGC GGG GCC TTC CAC CTT AAG GGC ACA GGA GAG CTG Leu Ser Phe Asp Gly Gly Ala Phe His Leu Lys Gly Thr Gly Glu Leu	115							

(2) INFORMATION FOR SEQ ID NO: 78:

CGG CTC TCA RGC GCA TTT CTC CGA GCA

Arg Leu Ser Xaa Ala Phe Leu Arg Ala

10 15

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 392 base pairs

(B) TYPE: NUCLEIC ACID

(C) STRANDEDNESS: DOUBLE

(D) TOPOLOGY: LINEAR

- (ii) MOLECULE TYPE: CDNA
- (vi) ORIGINAL SOURCE:
 - (A) ORGANISM: Homo Sapiens
 - (F) TISSUE TYPE: Lung (cells)
- (ix) FEATURE:
 - (A) NAME/KEY: sig_peptide
 - (B) LOCATION: 270..359
 - (C) IDENTIFICATION METHOD: Von Heijne matrix
 - (D) OTHER INFORMATION: score 5.7

seq LGVGCHFFHLALG/RF

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 78:

CAACTCTAAA AATCTTTACA CAGCCTATGG AGAGAAACCT AAGCAAGTSC CCAAATCCTC AAGTGTTCCA GGAGAATGAA AAGAATTATA AAGGCTAAAG AATTATTAGA ATCAGTGCAG TGAACATGTA GACCAAAGCA TTCCTGCATG CCAGGAAAAT GGTGCATTTG AATGTTTTTG CTTCTCATGA GAAAGGCAAT TTAAGTTGCA AGGCAAAGGC AAACTTTTGA AGATGGCCAC 240 GTGGACTCTG GATCCTTCTC TCTTTGTTC ATG ATT TGT TCT CCC TTC AGT GGT 293 Met Ile Cys Ser Pro Phe Ser Gly -30 TIT GCT CCT TGC CAA GCA TTA GGT ACC CTT GGG GTG GGA TGC CAC TTT Phe Ala Pro Cys Gln Ala Leu Gly Thr Leu Gly Val Gly Cys His Phe -15 TTC CAC TTA GCC TTG GGC AGG TTT CTT CTC TCC TTA TCC AAT AAT ATT Phe His Leu Ala Leu Gly Arg Phe Leu Leu Ser Leu Ser Asn Asn Ile 1 TAC 392 Tyr

- (2) INFORMATION FOR SEQ ID NO: 79:
 - (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 511 base pairs
 - (B) TYPE: NUCLEIC ACID
 - (C) STRANDEDNESS: DOUBLE
 - (D) TOPOLOGY: LINEAR
 - (ii) MOLECULE TYPE: CDNA
 - (vi) ORIGINAL SOURCE:
 - (A) ORGANISM: Homo Sapiens
 - (F) TISSUE TYPE: Pancreas
 - (ix) FEATURE:

(A)	NAME/KEY:	sig_pepti	ide		
(B)	LOCATION:	62355			
(C)	IDENTIFICA	ATION METH	HOD: Vo	n Heijne	matrix
(D)	OTHER INFO	ORMATION:	score	5.6	
			seq I	LLLRGADRY	/LQA/HI

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 79:

AAGGCAGTCC CGGCGCGGGG CAACCACAGC GCTCAGGAGG CCCTGAGTCC GGCACCTGGC											60					
				co GI					eu Xa					lu Va	TC TTC	109
					CAT His											157
CCA Pro	GCC Ala -65	GCC Ala	CCA Pro	GAT Asp	CCC Pro	TCG Ser -60	GAT Asp	CCC Pro	TGG Trp	GGC	AGA Arg -55	GAG Glu	TGC Cys	CTG Leu	CGG Arg	205
					CAC His -45											253
GAG Glu	GAA Glu	AGC Ser	CTG Leu	CAC His -30	TCA Ser	CTG Leu	CAG Gln	GAG Glu	AGG Arg -25	CTG Leu	CGT Arg	TAC Tyr	CCG Pro	GAC Asp -20	TCC Ser	301
					CTG Leu											349
					TAC Tyr											397
					GHA Xaa 20											445
					CGG Arg											493
				TCG Ser												511

(2) INFORMATION FOR SEQ ID NO: 80:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 340 base pairs
 - (B) TYPE: NUCLEIC ACID
 - (C) STRANDEDNESS: DOUBLE
 - (D) TOPOLOGY: LINEAR

•	
(ii) MOLECULE TYPE: CDNA	
(vi) ORIGINAL SOURCE:(A) ORGANISM: Homo Sapiens(F) TI5SUE TYPE: Thyroid	
<pre>(ix) FEATURE: (A) NAME/KEY: sig_peptide (B) LOCATION: 131265 (C) IDENTIFICATION METHOD: Von Heijne matrix (D) OTHER INFORMATION: score 5.6</pre>	
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 80:	
CTTTTCTGGT CCAGAGCCCC CACAGGAAGG AGAVNNBSAA AAGATTACTG TTGACTGCAC	60
AGTTCCAGGG GCCAAGGCAG AAGAAATCCT GGAGAAAGGT CCAAAGACAG CGATCTCCTG	120
GACATGAAAG ATG GAC AAG CTG ATA CCC AGC TTG AGC TCT CAA GAG AAC Met Asp Lys Leu Ile Pro Ser Leu Ser Ser Gln Glu Asn -45 -40 -35	169
AGA AAG GCG TCT CAC ACT CTC CAC AAA GCT AGA AAC AAA CAA CAC TGT Arg Lys Ala Ser His Thr Leu His Lys Ala Arg Asn Lys Gln His Cys -30 -25	217
GGA GGA TTT TTA CTG GTC ATA CAT TGG GTC ATG TGC CCT TCC CTG AGC Gly Gly Phe Leu Leu Val Ile His Trp Val Met Cys Pro Ser Leu Ser -15 -5	265
CAA TCT GCA GTC AGA AGG ATG AAG TAC TCT AAT TGG CCA GTT TTG GGT Gln Ser Ala Val Arg Arg Met Lys Tyr Ser Asn Trp Pro Val Leu Gly 1 5 15	313
CAC GTG CCT GTT CCT GGC TGT CAT TGC His Val Pro Val Pro Gly Cys His Cys 20 25	340
(2) INFORMATION FOR SEQ ID NO: 81:	
(i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 231 base pairs (B) TYPE: NUCLEIC ACID (C) STRANDEDNESS: DOUBLE (D) TOPOLOGY: LINEAR	
(ii) MOLECULE TYPE: CDNA	
<pre>(vi) ORIGINAL SOURCE: (A) ORGANISM: Homo Sapiens (F) TISSUE TYPE: Lung (cells)</pre>	
<pre>(ix) FEATURE: (A) NAME/KEY: sig_peptide (B) LOCATION: 127177 (C) IDENTIFICATION METHOD: Von Heijne matrix (D) OTHER INFORMATION: score 5.5</pre>	

64

seq LPVVLASPPVGHG/LP

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 81:

TAATTGGGCA GCCAGGGTCT CCTGGAGCAT GGCGGTGACA TTCAGAGCCC CACAGCAACT 60

CCGGGCATCC CACCTCTGCC TAGGTGGGAT ACATCTTGAG CCTACGGCAG TCCCTCTGTC 120

GGTCTC ATG AGC CAK CTT CTT CCT GTG GTT CTT GCC TCA CCT CCA GTA

Met Ser Xaa Leu Leu Pro Val Val Leu Ala Ser Pro Pro Val

-15 -10 -5

GGC CAT GGG CTT CCC TCC CCA GTA CCT CTG TTA CAG GAC CCC TGC CCC
Gly His Gly Leu Pro Ser Pro Val Pro Leu Leu Gln Asp Pro Cys Pro

1 5 10

CTC CCT GCT GTC GGG

Leu Pro Ala Val Gly

15

(2) INFORMATION FOR SEQ ID NO: 82:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 128 base pairs
 - (B) TYPE: NUCLEIC ACID
 - (C) STRANDEDNESS: DOUBLE
 - (D) TOPOLOGY: LINEAR
- (ii) MOLECULE TYPE: CDNA
- (vi) ORIGINAL SOURCE:
 - (A) ORGANISM: Homo Sapiens
 - (F) TISSUE TYPE: Lung (cells)
- (ix) FEATURE:
 - (A) NAME/KEY: sig_peptide
 - (B) LOCATION: 54..92
 - (C) IDENTIFICATION METHOD: Von Heijne matrix
 - (D) OTHER INFORMATION: score 5.5

seq MVLLTMIARVADG/LP

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 82:

AGACGGCGCC AAGGGCCTTC CGGGCCAGTG TTGGATCCCT GTAGTTTGTG AAG ATG
Met

GTG TTG CTA ACA ATG ATC GCC CGA GTG GCG GAC GGG CTC CCG CTG GCC
Val Leu Leu Thr Met Ile Ala Arg Val Ala Asp Gly Leu Pro Leu Ala
-10 -5 1

GCC TCG ATG CAG GAG GAC GAA GAG
Ala Ser Met Gln Glu Asp Glu Glu
5

(2) INFORMATION FOR SEQ ID NO: 83:

65

	(i)	(A (B (C) LE) TY) SI	NGTH PE: RANI	ARAC NUCI DEDNE OGY:	08 ba LEIC ESS:	ase p ACII DOU	pairs D	3							
	(ii) MO	LECU	JLE T	YPE:	CD	NA.									
	(vi	(A) OF	RGANI		Homo		piens e int		ine						
	(ix	(B (C) NA) LC	ME/F CATI ENTI	FICA	qmoo 1017/	oleme N ME1	io	bla	istn ity 1 n 159	928	33				
		(B (C (D	LO (E	ME/F CATI CENTI CHER	INFO	32. ATION DRMA	.190 N ME: TION	THOD:	core eq LI	5.5 LELLI	EVPLI					
	•	,	•													
CACT	raaag	CT G	CCC1	CCT	CC TA	ACTG:	rtcc		ATG 1		lis :					52
	GCT Ala -45															100
GCC Ala -30	TTC Phe	TCC Ser	ATG Met	AAG Lys	GCT Ala -25	ATC Ile	CAT His	ATA Ile	ACC Thr	AGA Arg -20	ATA Ile	AGA Arg	AGT Ser	GGT Gly	CTT Leu -15	148
	CTC Leu															196
	TGG Trp															208
(2)	INFO	RMAT	CION	FOR	SEQ	ID	NO:	84:								
	(i)	(P (E	A) LI B) T' C) S'	engti YPE: Trani	HARAC H: 2 NUC DEDNI OGY:	60 b LEIC ESS:	ase ACI DOU	pair D	s							

(ii) MOLECULE TYPE: CDNA

(ix) FEATURE: (A) NAME/KEY: sig_peptide (B) LOCATION: 147221 (C) IDENTIFICATION METHOD: Von Heijne matrix (D) OTHER INFORMATION: score 5.4	<pre>(vi) ORIGINAL SOURCE: (A) ORGANISM: Homo Sapiens (F) TISSUE TYPE: Pancreas</pre>
ATTATTTAA ATATTATCAT GGACAGTTGT TATTGGAAAT GACAGTTCAC CATTGCAAAT CAATGCAGTG GCTACAAGTC CCCCGTGTCA CACAAAAATA GACGGGGTAG CTCTGACAGC ATGCATCCTK ATGCCCTGTT CTATTA ATG ATG CAC TGC ACC CCA TCA GGG TCT Met Met His Cys Thr Pro Ser Gly Ser -25 GCA GCT GTA TCA TTA CTC ACA GAG ACG GTT CTG CCT TTG GCT TTC CCT Ala Ala Val Ser Leu Leu Thr Glu Thr Val Leu Pro Leu Ala Phe Pro -15 -10 -5 GGT CCT CCA TGG CTA GGA ACA TCA TTT AAC AGG GKT TTG Gly Pro Pro Trp Leu Gly Thr Ser Phe Asn Arg Xaa Leu 1 5 (2) INFORMATION FOR SEQ ID NO: 85: (i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 316 base pairs (B) TYPE: NUCLEIC ACID (C) STRANDEDNESS: DOUBLE (D) TOPOLOGY: LINEAR (ii) MOLECULE TYPE: CDNA (vi) ORIGINAL SOURCE: (A) ORGANISM: Homo Sapiens (F) TISSUE TYPE: Lung (ix) FEATURE: (A) NAME/KEY: sig_peptide (B) LOCATION: 95178 (C) IDENTIFICATION METHOD: Von Heijne matrix (D) OTHER INFORMATION: score 5.3 seq RISCAFSLASSTA/RQ	(A) NAME/KEY: sig_peptide(B) LOCATION: 147221(C) IDENTIFICATION METHOD: Von Heijne matrix(D) OTHER INFORMATION: score 5.4
CAATGCAGTG GCTACAAGTC CCCCGTGTCA CACAAAAATA GACGGGGTAG CTCTGACAGC ATGCATCCTK ATGCCCTGTT CTATTA ATG ATG CAC TGC ACC CCA TCA GGG TCT Met Met His Cys Thr Pro Ser Gly Ser -25 -20 GCA GCT GTA TCA TTA CTC ACA GAG ACG GTT CTG CCT TTG GCT TTC CCT Ala Ala Val Ser Leu Leu Thr Glu Thr Val Leu Pro Leu Ala Phe Pro -15 -10 -5 GGT CCT CCA TGG CTA GGA ACA TCA TTT AAC AGG GKT TTG Gly Pro Pro Trp Leu Gly Thr Ser Phe Asn Arg Xaa Leu 1 5 10 (2) INFORMATION FOR SEQ ID NO: 85: (i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 316 base pairs (B) TYPE: NUCLEIC ACID (C) STRANDEDNESS: DOUBLE (D) TOPOLOGY: LINEAR (ii) MOLECULE TYPE: CDNA (vi) ORIGINAL SOURCE: (A) ORGANISM: Homo Sapiens (F) TISSUE TYPE: Lung (ix) FEATURE: (A) NAME/KEY: sig_peptide (B) LOCATION: 95178 (C) IDENTIFICATION METHOD: Von Heijne matrix (D) OTHER INFORMATION: score 5.3 seq RISCAFSLASSTA/RQ	(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 84:
ATGCATCCTK ATGCCCTGTT CTATTA ATG ATG CAC TGC ACC CCA TCA GGG TCT Met Met His Cys Thr Pro Ser Gly Ser -25 GCA GCT GTA TCA TTA CTC ACA GAG ACG GTT CTG CCT TTG GCT TTC CCT Ala Ala Val Ser Leu Leu Thr Glu Thr Val Leu Pro Leu Ala Phe Pro -15 -10 GGT CCT CCA TGG CTA GGA ACA TCA TTT AAC AGG GKT TTG Gly Pro Pro Trp Leu Gly Thr Ser Phe Asn Arg Xaa Leu 1 (2) INFORMATION FOR SEQ ID NO: 85: (i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 316 base pairs (B) TYPE: NUCLEIC ACID (C) STRANDEDNESS: DOUBLE (D) TOPOLOGY: LINEAR (ii) MOLECULE TYPE: CDNA (vi) ORIGINAL SOURCE: (A) ORGANISM: Homo Sapiens (F) TISSUE TYPE: Lung (ix) FEATURE: (A) NAME/KEY: sig_peptide (B) LOCATION: 95178 (C) IDENTIFICATION METHOD: Von Heijne matrix (D) OTHER INFORMATION: score 5.3 seq RISCAFSLASSTA/RQ	ATTATTTTAA ATATTATCAT GGACAGTTGT TATTGGAAAT GACAGTTCAC CATTGCAAAT 6
Met Met His Cys Thr Pro Ser Gly Ser -25 -20 GCA GCT GTA TCA TTA CTC ACA GAG ACG GTT CTG CCT TTG GCT TTC CCT Ala Ala Val Ser Leu Leu Thr Glu Thr Val Leu Pro Leu Ala Phe Pro -15 -10 -5 GGT CCT CCA TGG CTA GGA ACA TCA TTT AAC AGG GKT TTG Gly Pro Pro Trp Leu Gly Thr Ser Phe Asn Arg Xaa Leu 1 5 10 (2) INFORMATION FOR SEQ ID NO: 85: (i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 316 base pairs (B) TYPE: NUCLEIC ACID (C) STRANDEDNESS: DOUBLE (D) TOPOLOGY: LINEAR (ii) MOLECULE TYPE: CDNA (vi) ORIGINAL SOURCE: (A) ORGANISM: Homo Sapiens (F) TISSUE TYPE: Lung (ix) FEATURE: (A) NAME/KEY: sig_peptide (B) LOCATION: 95178 (C) IDENTIFICATION METHOD: Von Heijne matrix (D) OTHER INFORMATION: score 5.3 seq RISCAFSLASSTA/RQ	CAATGCAGTG GCTACAAGTC CCCCGTGTCA CACAAAAATA GACGGGGTAG CTCTGACAGC 12
Ala Ala Val Ser Leu Leu Thr Glu Thr Val Leu Pro Leu Ala Phe Pro -15 -10 -5 GGT CCT CCA TGG CTA GGA ACA TCA TTT AAC AGG GKT TTG Gly Pro Pro Trp Leu Gly Thr Ser Phe Asn Arg Xaa Leu 1 5 10 (2) INFORMATION FOR SEQ ID NO: 85: (i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 316 base pairs (B) TypE: NUCLEIC ACID (C) STRANDEDNESS: DOUBLE (D) TOPOLOGY: LINEAR (ii) MOLECULE TYPE: CDNA (vi) ORIGINAL SOURCE: (A) ORGANISM: Homo Sapiens (F) TISSUE TYPE: Lung (ix) FEATURE: (A) NAME/KEY: sig_peptide (B) LOCATION: 95178 (C) IDENTIFICATION METHOD: Von Heijne matrix (D) OTHER INFORMATION: score 5.3 seq RISCAFSLASSTA/RQ	Met Met His Cys Thr Pro Ser Gly Ser
Gly Pro Pro Trp Leu Gly Thr Ser Phe Asn Arg Xaa Leu 1 5 10 (2) INFORMATION FOR SEQ ID NO: 85: (i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 316 base pairs (B) TYPE: NUCLEIC ACID (C) STRANDEDNESS: DOUBLE (D) TOPOLOGY: LINEAR (ii) MOLECULE TYPE: CDNA (vi) ORIGINAL SOURCE: (A) ORGANISM: Homo Sapiens (F) TISSUE TYPE: Lung (ix) FEATURE: (A) NAME/KEY: sig_peptide (B) LOCATION: 95178 (C) IDENTIFICATION METHOD: Von Heijne matrix (D) OTHER INFORMATION: score 5.3 seq RISCAFSLASSTA/RQ	Ala Ala Val Ser Leu Leu Thr Glu Thr Val Leu Pro Leu Ala Phe Pro
(i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 316 base pairs (B) TYPE: NUCLEIC ACID (C) STRANDEDNESS: DOUBLE (D) TOPOLOGY: LINEAR (ii) MOLECULE TYPE: CDNA (vi) ORIGINAL SOURCE: (A) ORGANISM: Homo Sapiens (F) TISSUE TYPE: Lung (ix) FEATURE: (A) NAME/KEY: sig_peptide (B) LOCATION: 95178 (C) IDENTIFICATION METHOD: Von Heijne matrix (D) OTHER INFORMATION: score 5.3 seq RISCAFSLASSTA/RQ	Gly Pro Pro Trp Leu Gly Thr Ser Phe Asn Arg Xaa Leu
(C) IDENTIFICATION METHOD: Von Heijne matrix (D) OTHER INFORMATION: score 5.3 seq RISCAFSLASSTA/RQ	<pre>(i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 316 base pairs (B) TYPE: NUCLEIC ACID (C) STRANDEDNESS: DOUBLE (D) TOPOLOGY: LINEAR (ii) MOLECULE TYPE: CDNA (vi) ORIGINAL SOURCE: (A) ORGANISM: Homo Sapiens (F) TISSUE TYPE: Lung (ix) FEATURE: (A) NAME/KEY: sig_peptide</pre>
	<pre>(C) IDENTIFICATION METHOD: Von Heijne matrix (D) OTHER INFORMATION: score 5.3</pre>

		~
-	b	1

GGAGGCGACA TGGAGACAGG GGCGGCCAAG CTGT ATG ACC AGG CCC TTT TG Met Thr Arg Pro Phe Tr -25									
TCC TGC AGC ACG TGG GCA ACG TCC AGG ATT TCC TGC GCG TTC TCT Ser Cys Ser Thr Trp Ala Thr Ser Arg Ile Ser Cys Ala Phe Ser -20 -15 -10	TTG 163 Leu								
GCT TCC TCT ACC GCA AGA CAG ACT TCT ATC GCT TGC TGC GCC ACC Ala Ser Ser Thr Ala Arg Gln Thr Ser Ile Ala Cys Cys Ala Thr -5 1 5 10	CAT 211 His								
CGG ACC GCA TGG GCT TCC CGC CCG GGG CCG CGC AGG CCT TGG TGC Arg Thr Ala Trp Ala Ser Arg Pro Gly Pro Arg Arg Pro Trp Cys 15 20 25	TGC 259 Cys								
AGG TAT TCA AAA CCT TTG ACC ACA TGG CCC GTC AGG ATG AGA Arg Tyr Ser Lys Pro Leu Thr Thr Trp Pro Val Arg Met Met Arg 30 35 40	AGA 307 Arg								
GAA GGC CTA Glu Gly Leu 45	316								
(2) INFORMATION FOR SEQ ID NO: 86: (i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 452 base pairs (B) TYPE: NUCLEIC ACID (C) STRANDEDNESS: DOUBLE (D) TOPOLOGY: LINEAR (ii) MOLECULE TYPE: CDNA (vi) ORIGINAL SOURCE: (A) ORGANISM: Homo Sapiens (F) TISSUE TYPE: Lung (cells) (ix) FEATURE: (A) NAME/KEY: sig_peptide (B) LOCATION: 192269 (C) IDENTIFICATION METHOD: Von Heijne matrix (D) OTHER INFORMATION: score 5.2 seq SCCLIQWPELSFS/NT (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 86:									
ACTAAACAGM CDNHWACATT CAGGTGGGKA CCTTCCTCCA GAAGTTACTC CTACA	ATCAGA 60								
AAAGATGACC CCCTTCTTCC TGATGCTCAG AGCCATCAAT ACTCCTTTGT CTCTC									
GCACCCAATT CATCTGGAAA TCCTATTGGC ACCACCTTAA ATATGTAAAC AGAAC									
CAAGAGCCCC C ATG GTC ACC CAC CTA ATT AGG GGG GTG GTG TTG CAC Met Val Thr His Leu Ile Arg Gly Val Val Leu Glr -25 -20 -19	n Gly								
TCC TGC TGT CTC ATA CAG TGG CCA GAG TTA AGC TTT TCA AAC ACA	AAT 278								

PCT/IB98/01233 WO 99/06439

									68							
Şer	Cys	Cys	Leu -10	Ile	Gln	Trp	Pro	Glu -5	Leu	Ser	Phe	Ser	Asn 1	Thr	Asn	
GGT Gly	GTT Val 5	TGT Cys	CCC Pro	ATC Ile	TAT Tyr	CCC Pro 10	CCA Pro	CCT Pro	TCT Ser	ATC Ile	ASG Xaa 15	TSC Xaa	CTG Leu	AGA Arg	ATG Met	326
TCA Ser 20	TCC Ser	TGC Cys	TCT Ser	CCT Pro	CTG Leu 25	ACT Thr	GTA Val	TCT Ser	CTC Leu	TGC Cys 30	CCT Pro	TGC Cys	TAT Tyr	GTA Val	GAA Glu 35	374
TGT Cys	GCA Ala	TCC Ser	ACC Thr	CCA Pro 40	GGG Gly	CCT Pro	CTC Leu	TGC Cys	TTG Leu 45	CTC Leu	TTT Phe	TCT Ser	TGG Trp	CCA Pro 50	AGA Arg	422
	ACC Thr															452
(2)	INFO	RMAT	ION	FOR	SEQ	ID N	10: 8	37:								
	(ii)	(A) LE B) TY C) ST O) TO OLECU RIGIN A) OF	INGTHE PE: PRANTOPOLO POLO POLO POLO POLO POLO POLO POL	HARACH: 25 NUCI DEDNE DGY: TYPE: GOURG ISM: E TYPE	4 ba EIC SS: LINE CDN	ASE FACILIES DOUB DOUB EAR NA	oairs D BLE	5							
		() () ()	A) NA B) LO C) II D) OT	ME/I CATI CENTI CHER	KEY: ION: IFICA INFO	60. ATION CRMA	.116 N ME: TION	THOD : so	core eq L	5.2 LVAFI	RVFL					
					DESCI											
ACT	GTAG	CCA A	\ATT <i>I</i>	\AAA!	rc To	SAGT	CTGG	C GT	rtcc'	VTTT	GGG	AAGG'	rgc :	rcag'	TAGCT	59
	TAC Tyr															107
	TTC Phe															155
	GGA Gly 15															203
	CCT Pro															251

		69		
30	35	40		45
CGG Arg				254
(2)	INFORMATION FOR SEQ ID NO: 88:			
	(i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 212 base pairs (B) TYPE: NUCLEIC ACID (C) STRANDEDNESS: DOUBLE (D) TOPOLOGY: LINEAR			
	(ii) MOLECULE TYPE: CDNA			
	<pre>(vi) ORIGINAL SOURCE: (A) ORGANISM: Homo Sapiens (F) TISSUE TYPE: Lung (cell.)</pre>	s)		
		Von Heijne re 5.2 FLFGLYSFRA		
	(xi) SEQUENCE DESCRIPTION: SEQ I	D NO: 88:		
	AAAAGCA AGAACTGGGA CTAGTTTCAG GATT			
AGT	GCTAGA TGAGCTA ATG TCA TCT AGA AA Met Ser Ser Arg As -20			
CTC Leu	TTT GGG CTT TAT TCT TTC AGG GCT G Phe Gly Leu Tyr Ser Phe Arg Ala V -10 -5	TT GAC AGT al Asp Ser	TCT AGG ATT Ser Arg Ile 1	AAG 158 Lys
CTC Leu 5	AGT TTA TTA ACC AAG GAG GAA GAA A Ser Leu Leu Thr Lys Glu Glu Glu T 10	CC CCT TCT hr Pro Ser 15	GCC TAC TAC Ala Tyr Tyr	AGA 206 Arg 20
	CTG Leu			212
(2)	INFORMATION FOR SEQ ID NO: 89:			
	(i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 140 base pairs (B) TYPE: NUCLEIC ACID (C) STRANDEDNESS: DOUBLE (D) TOPOLOGY: LINEAR			
	(ii) MOLECULE TYPE: CDNA			

(vi) ORIGINAL SOURCE:

PCT/IB98/01233

WO 99/06439

PCT/IB98/01233

70

(A) ORGANISM: Homo Sapiens (F) TISSUE TYPE: Lung (cells)	
<pre>(ix) FEATURE: (A) NAME/KEY: sig_peptide (B) LOCATION: 3386 (C) IDENTIFICATION METHOD: Von Heijne matrix (D) OTHER INFORMATION: score 5.1</pre>	
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 89:	
ATTTCAGAAA GGAGGAGTTG ATGATCTTCT AG ATG TAT ATG AAC ACC TGT CTA Met Tyr Met Asn Thr Cys Leu -15	53
TAT CTG CAT GTA TAT GTT TTG ACC TGC AGT GGT TGC AAT GTT GAT ATG Tyr Leu His Val Tyr Val Leu Thr Cys Ser Gly Cys Asn Val Asp Met -10 -5 1 5	101
TGT TCA AGA TTA TTC CTG TCT ACA AÄA CTG AAG GCC CGG Cys Ser Arg Leu Phe Leu Ser Thr Lys Leu Lys Ala Arg 10	140
(i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 221 base pairs (B) TYPE: NUCLEIC ACID (C) STRANDEDNESS: DOUBLE (D) TOPOLOGY: LINEAR (ii) MOLECULE TYPE: CDNA (vi) ORIGINAL SOURCE: (A) ORGANISM: Homo Sapiens (F) TISSUE TYPE: Lung (cells) (ix) FEATURE: (A) NAME/KEY: sig_peptide (B) LOCATION: 54200 (C) IDENTIFICATION METHOD: Von Heijne matrix (D) OTHER INFORMATION: score 5.1 seq VALSASLPQCSLG/LL (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 90:	
ACTCCACCCC CAGCCCTGCC CCTCCACACC TCGCCCCAAG AGCAGCCAGA GAG ATG Met	56
TCC TGC CGA CAA CCC ACC CCA ACA CAG TGT TCC CTA CTC CCA AAC GAC Ser Cys Arg Gln Pro Thr Pro Thr Gln Cys Ser Leu Leu Pro Asn Asp -45 -40 -35	104
AAC CGT GTC TCT ACG AGG GGA GGG GAC AGT GCT GGG CGC CAC CGC CAA Asn Arg Val Ser Thr Arg Gly Gly Asp Ser Ala Gly Arg His Arg Gln -30 -25 -20	152

GTC CCT CAG GTG GCT CTG AGT GCA AGT CTG CCC CAA TGC TCC CTT GGA Val Pro Gln Val Ala Leu Ser Ala Ser Leu Pro Gln Cys Ser Leu Gly -15 -5	200
CTC CTC ATA AAC CCC CGC CTG Leu Leu Ile Asn Pro Arg Leu 1 5	221
(2) INFORMATION FOR SEQ ID NO: 91: (i) SEQUENCE CHARACTERISTICS:	
(A) LENGTH: 158 base pairs (B) TYPE: NUCLEIC ACID (C) STRANDEDNESS: DOUBLE (D) TOPOLOGY: LINEAR	
(ii) MOLECULE TYPE: CDNA	
(vi) ORIGINAL SOURCE:(A) ORGANISM: Homo Sapiens(F) TISSUE TYPE: Thyroid	
<pre>(ix) FEATURE: (A) NAME/KEY: sig_peptide (B) LOCATION: 84143 (C) IDENTIFICATION METHOD: Von Heijne matrix (D) OTHER INFORMATION: score 5.1</pre>	
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 91:	
GAAATGTTTC CTTTTGTGTT AAAAAAGGTG AAGTTTTGGG ATTACTAGGA CACAATGGAG	60
	60
CTGGTAAAAG TACTTCVATT AAA ATG ATA ACT GGG TGC ACA AAG CCA ACT GCA Met Ile Thr Gly Cys Thr Lys Pro Thr Ala -20 -15	113
GGA GTG GTG TTA CAA GGC AGC AGA GCA TCA GTA AGG CAA CGG Gly Val Val Leu Gln Gly Ser Arg Ala Ser Val Arg Gln Arg -10 -5 1 5	158
(2) INFORMATION FOR SEQ ID NO: 92:	
(i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 70 base pairs (B) TYPE: NUCLEIC ACID (C) STRANDEDNESS: DOUBLE (D) TOPOLOGY: LINEAR	
(ii) MOLECULE TYPE: CDNA	
<pre>(vi) ORIGINAL SOURCE: (A) ORGANISM: Homo Sapiens (F) TISSUE TYPE: Lung (cells)</pre>	

218

WG 99/00-037		I CI/ID/0/012
	72	•
(ix) FEATURE: (A) NAME/KEY: sig_pering (B) LOCATION: 2061 (C) IDENTIFICATION MINER (D) OTHER INFORMATION	ETHOD: Von Heijne matrix	
(xi) SEQUENCE DESCRIPTION	: SEQ ID NO: 92:	
TATCTAGTGT GTCACGCCA ATG GGC TMet Gly Lo	TG GAC TTA ATC CTT TCT TTC eu Asp Leu Ile Leu Ser Phe -10	
TCT TCC CCC GGT CCT GGG Ser Ser Pro Gly Pro Gly 1		70
(2) INFORMATION FOR SEQ ID NO:	93:	
(i) SEQUENCE CHARACTERIST (A) LENGTH: 218 base (B) TYPE: NUCLEIC AC (C) STRANDEDNESS: DO (D) TOPOLOGY: LINEAR	pairs ID UBLE	
(ii) MOLECULE TYPE: CDNA		
(vi) ORIGINAL SOURCE: (A) ORGANISM: Homo S (F) TISSUE TYPE: Col		
(ix) FEATURE: (A) NAME/KEY: sig_pe (B) LOCATION: 1219 (C) IDENTIFICATION M (D) OTHER INFORMATIO	7 ETHOD: Von Heijne matrix	
(xi) SEQUENCE DESCRIPTION	: SEQ ID NO: 93:	
GAATTCCAGG G ATG CGG GAG GAC A Met Arg Glu Asp A -60	AT GAG CAT GAA CGA AAC GTG sn Glu His Glu Arg Asn Val -55	
GGA GTT GAG AAC GTA AAG GAA GA Gly Val Glu Asn Val Lys Glu Gl -45	u Gly Gly Asp Glu Asp Leu S	
GGA GAT GAG GGC TGC CAA GTC CT	'A AGA CAC AGG CTC AGG GTC 1	rgc Agg 146

Gly Asp Glu Gly Cys Gln Val Leu Arg His Arg Leu Arg Val Cys Arg

AAG GTC GGC TTG TTG GAT CGT CTC TGT GCG CTG ACT TCT CTT TGC TCC

Lys Val Gly Leu Leu Asp Arg Leu Cys Ala Leu Thr Ser Leu Cys Ser

-30

CCA GGG CCT CTA CCC GCT ACC CTG

-15

73

Pro Gly Pro Leu Pro Ala Thr Leu

(2) INFORMATION FOR SEQ ID NO: 94:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 96 base pairs
 - (B) TYPE: NUCLEIC ACID
 - (C) STRANDEDNESS: DOUBLE
 - (D) TOPOLOGY: LINEAR
- (ii) MOLECULE TYPE: CDNA
- (vi) ORIGINAL SOURCE:
 - (A) ORGANISM: Homo Sapiens
 - (F) TISSUE TYPE: Colon
- (ix) FEATURE:
 - (A) NAME/KEY: sig_peptide
 - (B) LOCATION: 10..63
 - (C) IDENTIFICATION METHOD: Von Heijne matrix
 - (D) OTHER INFORMATION: score 4.9

seq GAVVSSWAXCSLG/XP

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 94:

AGCCTTCGC ATG GGG AAA CGG GCT GGT GCA GTG GTG TCA TCT TGG GCT CAN 51

Met Gly Lys Arg Ala Gly Ala Val Val Ser Ser Trp Ala Xaa

-15

-10

-5

TGC AGC CTC GGA SKT CCT GGG ATC CAG CGA TCC TCC CGC TTA ACG
Cys Ser Leu Gly Xaa Pro Gly Ile Gln Arg Ser Ser Arg Leu Thr
1 5 10

- (2) INFORMATION FOR SEQ ID NO: 95:
 - (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 302 base pairs
 - (B) TYPE: NUCLEIC ACID
 - (C) STRANDEDNESS: DOUBLE
 - (D) TOPOLOGY: LINEAR
 - (ii) MOLECULE TYPE: CDNA
 - (vi) ORIGINAL SOURCE:
 - (A) ORGANISM: Homo Sapiens
 - (F) TISSUE TYPE: Lung (cells)
 - (ix) FEATURE:
 - (A) NAME/KEY: sig_peptide
 - (B) LOCATION: 57..119
 - (C) IDENTIFICATION METHOD: Von Heijne matrix
 - (D) OTHER INFORMATION: score 4.9

seq WLLSDILGQGATA/NV

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 95:

AGTGGCGCG CGGAGACCCG GCTGGTATAA CAAGAGGATT	CCCMCLMCCL CCCLLC LTC								
AGIGGCGCG CGGAGACCCG GCIGGIAIAA CAAGAGGAIT	GCCTGATCCA GCCAAG ATG 59 Met								
CAG AGC ACT TCT AAT CAT CTG TGG CTT TTA TCT Gln Ser Thr Ser Asn His Leu Trp Leu Leu Ser -10	GAT ATT TTA GGC CAA 107 Asp Ile Leu Gly Gln -5								
GGA GCT ACT GCA AAC GTC TTT CGT GGA AGA CAT Gly Ala Thr Ala Asn Val Phe Arg Gly Arg His	AAG AAA ACT GGT GAT 155 Lys Lys Thr Gly Asp 10								
TTA TTT GCT ATC AAA GTA TTT AAT AAC ATA AGC Leu Phe Ala Ile Lys Val Phe Asn Asn Ile Ser 15 20	TTC CTT CGT CCA GTG 203 Phe Leu Arg Pro Val 25								
GAT GTT CAA ATG AGA GAA TTT GAA GTG TTG AAA Asp Val Gln Met Arg Glu Phe Glu Val Leu Lys 30	AAA CTC AAT CAC AAA 251 Lys Leu Asn His Lys 40								
AAT ATT GTC AAA TTA TTT GCT ATT GAA GAG GAG Asn Ile Val Lys Leu Phe Ala Ile Glu Glu Glu 45 50 55	ACA ACA ACA AGA CGG 299 Thr Thr Thr Arg Arg 60								
CGG Arg	302								
(2) INFORMATION FOR SEQ ID NO: 96: (i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 309 base pairs (B) TYPE: NUCLEIC ACID (C) STRANDEDNESS: DOUBLE (D) TOPOLOGY: LINEAR (ii) MOLECULE TYPE: CDNA (vi) ORIGINAL SOURCE: (A) ORGANISM: Homo Sapiens (D) DEVELOPMENTAL STAGE: Fetal (F) TISSUE TYPE: liver (ix) FEATURE: (A) NAME/KEY: sig_peptide (B) LOCATION: 109255 (C) IDENTIFICATION METHOD: Von Heijne matrix (D) OTHER INFORMATION: score 4.9 seq LLCLSGLELEPSA/SD (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 96:									
AAAGTCACWW AAACAAGTTA GTAAGTTATG AAATTAGTTA	CATGCATGGA ATATTCATGC 60								
TTGAATGARS CTTAAAAGTT TATCTAACCC AAACCTCAGA	TTTTACAG ATG AAG AAG 117 Met Lys Lys								
CTG AGG CCC AGC CAG GAA CAG CTG AAC TGT CCG	GAG CCA CAA CTG GCA 165								

Leu Arg Pro Ser Gln Glu Gln Leu Asn Cys Pro Glu Pro Gln Leu Ala -45 -40 -35	
GAT GGC AGA GCT GGG ATT AGA TTG CTA GTA ACC TGG CTC CAA CCT GCA Asp Gly Arg Ala Gly Ile Arg Leu Leu Val Thr Trp Leu Gln Pro Ala -30 -25 -20 -15	213
CCT CTG CTC TGC CTC TCT GGG CTG GAA CTA GAA CCC AGT GCT TCT GAC Pro Leu Cys Leu Ser Gly Leu Glu Leu Glu Pro Ser Ala Ser Asp -10 -5 1	261
TTT GGA TTT AGT TCT CAT ACC ACT CTC CTG TGC TGC CTT GTT GAA AAT Phe Gly Phe Ser Ser His Thr Thr Leu Leu Cys Cys Leu Val Glu Asn 5 10 15	309
(2) INFORMATION FOR SEQ ID NO: 97: (i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 372 base pairs (B) TYPE: NUCLEIC ACID (C) STRANDEDNESS: DOUBLE (D) TOPOLOGY: LINEAR (ii) MOLECULE TYPE: CDNA (vi) ORIGINAL SOURCE: (A) ORGANISM: Homo Sapiens (F) TISSUE TYPE: Large intestine (ix) FEATURE: (A) NAME/KEY: sig_peptide (B) LOCATION: 151207 (C) IDENTIFICATION METHOD: Von Heijne matrix (D) OTHER INFORMATION: score 4.9 seq RLLFWSIFSSVTC/RK (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 97:	
AGATTGGTCG AACAAACCAG TATTATGCAA ACCTCATCCA AACCCTCTGA TTTCCTTAAC	60
TTGGCTAAGA AAAAGAGGAA GTTCTCCGAG TGACTCACCA CTGTGGTTCT ACTATGCCTT	120
CTGACCCCGT CTTGGACTTC AACTGGGAGA ATG TGG AGC CAT TTG AAC AGG CTC Met Trp Ser His Leu Asn Arg Leu -15	174
CTC TTC TGG AGC ATA TTT TCT TCT GTC ACT TGT AGA AAA GCT GTA TTG Leu Phe Trp Ser Ile Phe Ser Ser Val Thr Cys Arg Lys Ala Val Leu -10 -5 1 5	222
GAT TGT GAG GCA ATG AAA ACA AAT GAA TTC CCT TCT CCA TGT TTG GAC Asp Cys Glu Ala Met Lys Thr Asn Glu Phe Pro Ser Pro Cys Leu Asp 10 15 20	270
TCA AAG ACT AAG GTG GTT ATG AAG GGT CAA AAT GTA TCT ATG TTT TGT Ser Lys Thr Lys Val Val Met Lys Gly Gln Asn Val Ser Met Phe Cys 25 30 35	318

WO 99/06439 PCT/IB98/01233 76 TCC CAT AAG AAC AAA TCA CTG CAG ATC ACC TAT TCA TTG TTT CGA CGT Ser His Lys Asn Lys Ser Leu Gln Ile Thr Tyr Ser Leu Phe Arg Arg 40 AAG ACA 372 Lys Thr 55 (2) INFORMATION FOR SEQ ID NO: 98: (i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 384 base pairs (B) TYPE: NUCLEIC ACID (C) STRANDEDNESS: DOUBLE (D) TOPOLOGY: LINEAR (ii) MOLECULE TYPE: CDNA (vi) ORIGINAL SOURCE: (A) ORGANISM: Homo Sapiens (F) TISSUE TYPE: Lung (cells) (ix) FEATURE: (A) NAME/KEY: sig_peptide (B) LOCATION: 136..291 (C) IDENTIFICATION METHOD: Von Heijne matrix (D) OTHER INFORMATION: score 4.8 seq SLLLAQATSNVVC/SL (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 98: AATGGGCAGG GACCGAGGCT GAAAGAGACA GGAGTCCTCC AGCCGAGGAG AAGCTGATGC AGCTGGGCTT GTCCCAGAGG GAGAGGGGTT TTCTTCTCCA ACGGGGAGCG GTGGAGGCAG 120 CTGAGGAAGT TTACC ATG CTT GCT CTG CGG GAC CTG GGC ATG GGG AAG CGA 171 Met Leu Ala Leu Arg Asp Leu Gly Met Gly Lys Arg -50 -45 GAA GGC GAG GAG CTG ATC CAG GCG GAG GCC CGG TGT CTG GTG GAG ACA 219 Glu Gly Glu Glu Leu Ile Gln Ala Glu Ala Arg Cys Leu Val Glu Thr -40 -30 TTC CAG GGG ACA GAA GGA CGC CCA TTC GAT CCC TCC CTG CTG GCC 267 Phe Gln Gly Thr Glu Gly Arg Pro Phe Asp Pro Ser Leu Leu Leu Ala -20 CAG GCC ACC TCC AAC GTA GTC TGC TCC CTC TTT GGC CTC CGC TTC 315 Gln Ala Thr Ser Asn Val Val Cys Ser Leu Leu Phe Gly Leu Arg Phe -5 1 TCC TAT GAG GAT AAG GAG TTC CAG GCC GTG GTC CGG GCA GCT GGT GGT 363 Ser Tyr Glu Asp Lys Glu Phe Gln Ala Val Val Arg Ala Ala Gly Gly

15

394

10

25

ACC TGC TGG GAG TCA GCT CCC

Thr Cys Trp Glu Ser Ala Pro

(2)	INF	ORMA:	rion	FOR	SEQ	ID 1	10:	99:								
	(i)	I)	A) LI B) T' C) S'	ENGTI YPE : TRANI	H: 21 NUCI DEDNI		ase ACI DOU	pair: D	S							
	(i:	i) MC	DLEC	JLE 7	YPE:	: CDI	NA									
<pre>(vi) ORIGINAL SOURCE: (A) ORGANISM: Homo Sapiens (F) TISSUE TYPE: Liver</pre>																
<pre>(ix) FEATURE: (A) NAME/KEY: sig_peptide (B) LOCATION: 1773 (C) IDENTIFICATION METHOD: Von Heijne matrix (D) OTHER INFORMATION: score 4.8</pre>																
	(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 99:															
ACG	CAGC!	rgg 1	rcago	CA AT	rg Ci	rc ac	GT G er V	al G	GA GO Ly Al 15	CC AC	GC AC	CC AG	er Le	rG T(≥u C ₁ LO	GT GGG /s Gly	5 52 ,
TGC Cys	CTG Leu	CGC Arg -5	CAA Gln	CTG Leu	CGG Arg	TGC Cys	AGC Ser 1	ATG Met	CTG Leu	GAT Asp	CTG Leu 5	CAG Gln	TGG Trp	AGC Ser	TTT Phe	100
CTC Leu 10	GAG Glu	GAT Asp	GGG Gly	GAG Glu	CCA Pro 15	TGC Cys	AGA Arg	GCC Ala	CGC Arg	CTC Leu 20	TCA Ser	CCC Pro	CTG Leu	CCT Pro	CCA Pro 25	148
CTT Leu	GCT Ala	CAC His	TTG Leu	GCT Ala 30	GGA Gly	ATC Ile	TGG Trp	ATA Ile	GTC Val 35	CTG Leu	CCA Pro	AGG Arg	GCT Ala	AGT Ser 40	TTT Phe	196
		ATG Met														217
(2) INFORMATION FOR SEQ ID NO: 100: (i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 152 base pairs																
	(B) TYPE: NUCLEIC ACID (C) STRANDEDNESS: DOUBLE (D) TOPOLOGY: LINEAR															

(ii) MOLECULE TYPE: CDNA

(A) ORGANISM: Homo Sapiens

(vi) ORIGINAL SOURCE:

Met Thr Ser His -15

284

WO 99/06439	РСТ/ІВ	98/0123
	78	
(F) TISSUE TYPE: Lung	3	
(ix) FEATURE: (A) NAME/KEY: sig_per (B) LOCATION: 84140 (C) IDENTIFICATION ME (D) OTHER INFORMATION) ETHOD: Von Heijne matrix	
(xi) SEQUENCE DESCRIPTION:	SEQ ID NO: 100:	
ATTTCCTTTG GGTGCTCCAA CTCTTGTAA	AC TTCACAGGCA ACAACGTTAT CTACGTTCCA	60
	Phe Gln Gln Met Tyr Val Leu Leu Ser -15 -10	113
CAG TTT CTC TAC CCY TTA GCT TAT Gln Phe Leu Tyr Pro Leu Ala Tyr -5		152
(2) INFORMATION FOR SEQ ID NO:	101:	
(i) SEQUENCE CHARACTERISTI (A) LENGTH: 296 base (B) TYPE: NUCLEIC ACI (C) STRANDEDNESS: DOU (D) TOPOLOGY: LINEAR	pairs ID ·	
(ii) MOLECULE TYPE: CDNA		
(vi) ORIGINAL SOURCE: (A) ORGANISM: Homo Sa (F) TISSUE TYPE: Lung		
(ix) FEATURE: (A) NAME/KEY: sig_per (B) LOCATION: 22527 (C) IDENTIFICATION ME (D) OTHER INFORMATION	72 ETHOD: Von Heijne matrix	
(xi) SEQUENCE DESCRIPTION	: SEQ ID NO: 101:	
ATTCTAAATG TTCTCAGATG ACTTTCAT	TT CACTCACATC CTGAGATTAT GCTGTGGCAT	60
TAATCAATTA CTACTTTTTA TAATTTTT	TT TCAGTAGATC TACTAGAAAG AAAGAAAAGA	120
TCTACTAGAA GTCAAGAAAA TGAATGCTC	CA GTTAAGTGAG GGGAGAAGGA AAGTATTTAT	180
TGTTATTACT ATTTTTTATA ACGTCAGA	GA CCTAATAGTG AAGA ATG ACT AGT CAT	236

TTT TGT KTG ATT GGT TTT TTA TCT TAT ACA ACC TCC TTG GTC TAC TGG

Phe Cys Xaa Ile Gly Phe Leu Ser Tyr Thr Thr Ser Leu Val Tyr Tro

-10

AAT GCA GGC CGG Asn Ala Gly Arg 5													
(2) INFORMATION FOR SEQ ID NO: 102:													
(i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 255 base pairs (B) TYPE: NUCLEIC ACID (C) STRANDEDNESS: DOUBLE (D) TOPOLOGY: LINEAR													
(ii) MOLECULE TYPE: CDNA													
<pre>(vi) ORIGINAL SOURCE: (A) ORGANISM: Homo Sapiens (F) TISSUE TYPE: Thyroid</pre>													
<pre>(ix) FEATURE: (A) NAME/KEY: sig_peptide (B) LOCATION: 115189 (C) IDENTIFICATION METHOD: Von Heijne matrix (D) OTHER INFORMATION: score 4.8</pre>													
TOTOTO ACTO COTOTO CONTENT HOMOTO AND THE THORN CHOOSE AND													
TCTCTGACTT CCTCTGCTTT TCTCTGAATC TTCTACTCTG CTGCAGCCAC ATTGGCT													
TTCTTCAAGC ACTGGAAGCT CGCTCCTGTT TCAGGGCCTT TGTTCCAAAC ATAC AT	t												
ATT TGC TCT CTC ACT CCC TTC AGG TCT TTG ACT AAT GTC CTT CTC AG Ile Cys Ser Leu Thr Pro Phe Arg Ser Leu Thr Asn Val Leu Leu Se -20 -15 -10	T 165												
GGA AGC CTT CTC CGA TCA CTT TGT TTG AAA TAT AAA CCA CTC ACC TC Gly Ser Leu Leu Arg Ser Leu Cys Leu Lys Tyr Lys Pro Leu Thr Se	C 213												
ATC TTC CTT GTC TCA ATG TGT CCT ATA CCC TTT CCC TGC CAT Ile Phe Leu Val Ser Met Cys Pro Ile Pro Phe Pro Cys His 10 15 20	255												
(2) INFORMATION FOR SEQ ID NO: 103:													
(i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 194 base pairs (B) TYPE: NUCLEIC ACID (C) STRANDEDNESS: DOUBLE													

(D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: CDNA

80

<pre>(vi) ORIGINAL SOURCE: (A) ORGANISM: Homo Sapiens (F) TISSUE TYPE: Colon</pre>	
<pre>(ix) FEATURE: (A) NAME/KEY: sig_peptide (B) LOCATION: 93173 (C) IDENTIFICATION METHOD: Von Heijne matrix (D) OTHER INFORMATION: score 4.7</pre>	
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 103:	
ACTICIAGGG AAGGGCTGGG GAGGAGGGGG TGGTGACACG GTGGAGACAC CGGCTAGGCC	60
AGGGGGCCTG CCCTTGGGAC AGGTCCAGAC CC ATG GAG CCC CCC GGG AGG AGC Met Glu Pro Pro Gly Arg Ser -25	113
AGC AGC CTT CCC TTT TCC CCT CCC GCA CTC ACT CTC ACC TTC TTG CCC Ser Ser Leu Pro Phe Ser Pro Pro Ala Leu Thr Leu Thr Phe Leu Pro -15 -10 -5	161
CCA TCG CCC ACC CTG CCA CTT CCC TCC CCT GGG Pro Ser Pro Thr Leu Pro Leu Pro Ser Pro Gly 1 5	194
(2) INFORMATION FOR SEQ ID NO: 104:	
(i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 282 base pairs (B) TYPE: NUCLEIC ACID (C) STRANDEDNESS: DOUBLE (D) TOPOLOGY: LINEAR	
(ii) MOLECULE TYPE: CDNA	
<pre>(vi) ORIGINAL SOURCE: (A) ORGANISM: Homo Sapiens (F) TISSUE TYPE: Colon</pre>	
<pre>(ix) FEATURE: (A) NAME/KEY: sig_peptide (B) LOCATION: 94258 (C) IDENTIFICATION METHOD: Von Heijne matrix (D) OTHER INFORMATION: score 4.7</pre>	
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 104:	
AAGGGCGGCT GCCTAGCACC CGGAAGAGCC GTCAACTTAG CGAGCGCAAC AGGCTGCCGC	60
TGAGGAGCTG GAGCTGGTGG GGACTGGGCC GCA ATG GAC AAG CTG AAG AAG GTG Met Asp Lys Leu Lys Lys Val -55 -50	114
CTG AGC GGG CAG GAC ACG GAG GAC CGG AGC GGC CTG TCC GAG GTT GTT	162

									U	4						
Leu	Ser	Gly	Gln -45	Asp	Thr	Glu	Asp	Arg -40	Ser	Gly	Leu	Ser	Glu -35	Val	Val	
GAG Glu	GCA Ala	TCT Ser -30	TCA Ser	TTA Leu	AGC Ser	TGG Trp	AGT Ser -25	ACC Thr	AGG Arg	ATA Ile	AAA Lys	GGC Gly -20	TTC Phe	ATT Ile	GCG Ala	210
TGT Cys	TTT Phe -15	GCT Ala	ATA Ile	GGA Gly	ATT Ile	CTC Leu -10	TGC Cys	TCA Ser	CTG Leu	CTG Leu	GGT Gly -5	ACT Thr	GTT Val	CTG Leu	CTG Leu	258
					GGA Gly											282
(2)	(i) (ii (vi) SEQ (# (# (# (# (# (# (# (# (# (# (# (# (#	QUENC A) LE B) TY C) ST O) TO OLECT RIGIN A) OE EATUE A) NA B) LO C) II	CE CHENGTHE PROPOSE OF COLUMN AL SECOND COLUMN AL SECOND COLUMN ACCORDANCE OF COLUMN ACCORDAN	SEQ HARACH H: 17 NUCI DEDNE DGY: TYPE: SOURCE ISM: E TYPE LIFICA INFO	CTERION DRMAT	ISTICASE FACILI DOUBLEAR NA D SapLung _pept .158 N MET	CS: pairs Diens (cel cide	S lls) : Vor core eq LO	4.7 SMVCI	FSL					
ATTO	CTCTT	ר דד ז	GTT	rtgti	TT CT	TTTG	CTTTI	CAC	CATGI	TAC	CAGO	CTACA	ATA A	\TTT(CTTGAC	60
AGA <i>I</i>	\AAA <i>t</i>	AAT P	\AAT <i>i</i>	LAAT <i>F</i>	AG TO	Me	rg TA et Ty 25	AC TO	CC AC	GG CA	AT AC S Th	ir Va	TA AF	AA Ci ys Le	TA AAA eu Lys	113
CAA Gln -15	GGT Gly	TTG Leu	GGT Gly	ATG Met	GTT Val -10	TGT Cys	ATT Ile	TTC Phe	AGT Ser	TTA Leu -5	AGG Arg	CTG Leu	CAA Gln	GCA Ala	GTA Val 1	161
	ACA Thr															179

- (2) INFORMATION FOR SEQ ID NO: 106:
 - (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 212 base pairs

									82							
		(B) T' C) S' D) T(TRANI	DEDNI	ESS:	DOU									
	(i:	i) M	OLEC	ULE '	TYPE	: CDI	NA									
	(v:	(RIGII A) OI F) T	RGAN:	ISM:	Homo										
	<pre>(ix) FEATURE: (A) NAME/KEY: sig_peptide (B) LOCATION: 48119 (C) IDENTIFICATION METHOD: Von Heijne matrix (D) OTHER INFORMATION: score 4.6</pre>															
	(x:	i) Si	EQUEI	NCE I	DESC	RIPT	ION:	SEQ	ID I	NO: 3	106:					
ATA	GCAA'	rag i	AAAG!	AGCC!	AG AA	ATATO	GTGC'	r tt <i>i</i>	AGTT:	rgtt	TTA	ATGA		TAT Tyr		56
TCA Ser	CTT Leu -20	TTA Leu	GTT Val	GAC Asp	TAT Tyr	TTC Phe -15	CCT Pro	AGT Ser	CTT Leu	TTG Leu	CTA Leu -10	TAC Tyr	AGT Ser	TTA Leu	CCC Pro	104
TTG Leu -5	AAC Asn	ATC Ile	ATA Ile	GGT Gly	TTG Leu 1	AAC Asn	TGT Cys	GCA Ala	TAT Tyr 5	CCA Pro	CTT Leu	ATA Ile	AAT Asn	AAT Asn 10	TTT Phe	152
TTA Leu	AAA Lys	AAT Asn	AAT Asn 15	AGT Ser	TAC Tyr	ACT Thr	TGT Cys	GTG Val 20	CVT Xaa	GTC Val	CCT Pro	CTT Leu	GCT Ala 25	TTC Phe	CCT Pro	200
	ATG Met															212
(2)	INFO	ORMA'	TION	FOR	SEQ	ID I	NO: :	107:			٠					
	(i)	() ()	QUENC A) LI B) T' C) S' D) To	engti YPE : Irani	H: 34 NUCI DEDNI	45 ba LEIC ESS:	ase p ACII	pair: D	5							
	(i:	i) M	OLEC	ULE :	TYPE	: CDI	A									
	(v:	(.	RIGII A) O F) T	RGAN:	ISM:	Home										
	(i:	(.	EATU A) N B) L	AME/	KEY:	sig 109	_pep	tide 9								

(C) IDENTIFICATION METHOD: Von Heijne matrix (D) OTHER INFORMATION: score 4.5

seq LPTQFLFLLGVLG/IF

4	(xi)	SEQUENCE	DESCRIPTION:	SEO	TD NO.	107
3	(A + /	200001100	DESCRIPTION.	360	ID NO:	111/

ATTAGCTSTC CAAGGTCTCC CCCAGCACTG AGGAGCTCGC CTGCTGCCCT CTTGCGCGCG 60
GGAAGCAGCA CCAAGTTCAC GGCCAACGCC TTGGCACTAG GGTCCAGA ATG GCT ACA 117 Met Ala Thr -75
ACA GTC CCT GAT GGT TGC CGC AAT GGC CTG AAA TCC AAG TAC TAC AGA 165 Thr Val Pro Asp Gly Cys Arg Asn Gly Leu Lys Ser Lys Tyr Tyr Arg -70 -65 -60
CTT TGT GAT AAG GCT GAA GCT TGG GGC ATC GTC CTA GAA ACG GTG GCC 213 Leu Cys Asp Lys Ala Glu Ala Trp Gly Ile Val Leu Glu Thr Val Ala -55 -50 -45
ACA GCC GGG GTT GTG ACC TCG GTG GCC TTC ATG CKG ACT CTC CCG ATC Thr Ala Gly Val Val Thr Ser Val Ala Phe Met Xaa Thr Leu Pro Ile -40 -35 -30
CTC GTC TGC AAG GTG CAG GAC TCC AAC AGG CGA AAA ATG CTG CCT ACT Seu Val Cys Lys Val Gln Asp Ser Asn Arg Arg Lys Met Leu Pro Thr -25 -20 -15
CAG TTT CTC TTC CTC CTG GGT GTG TTG GGC ATC TTT Sin Phe Leu Phe Leu Leu Gly Val Leu Gly Ile Phe -10 -5 1
(2) INFORMATION FOR SEQ ID NO: 108: (i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 266 base pairs (B) TYPE: NUCLEIC ACID (C) STRANDEDNESS: DOUBLE (D) TOPOLOGY: LINEAR (ii) MOLECULE TYPE: CDNA (vi) ORIGINAL SOURCE: (A) ORGANISM: Homo Sapiens (F) TISSUE TYPE: Large intestine (ix) FEATURE: (A) NAME/KEY: sig_peptide (B) LOCATION: 63113 (C) IDENTIFICATION METHOD: Von Heijne matrix (D) OTHER INFORMATION: SCORE 4.5 seq HLDHLFFSGVVLG/QG
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 108: AACTGGCATC ACTGCTAATG AAAAATGCTG GACTTGGGAT GATATTATGM AGTTAGAAAT 60
IC PEC BOD MEG CAG CAG CHE CAM CAM MED TO THE TAX
Met Arg Leu Gln His Leu Asp His Leu Phe Phe Ser Gly Val Val -15 -10 -5

PCT/IB98/01233

CTG Leu	GGG Gly	CAG Gln 1	GGG Gly	TTG Leu	GAC Asp	CTT Leu 5	GGA Gly	AGA Arg	GTG Val	TGT Cys	TTA Leu 10	CGA Arg	AAA Lys	TGG Trp	GGT Gly	155
TAC Tyr 15	AGA Arg	AGA Arg	TGT Cys	GAA Glu	GAT Asp 20	ATT Ile	TGT Cys	TGG Trp	ATT Ile	AAA Lys 25	ACC Thr	AAT Asn	AAA Lys	AAC Asn	AAT Asn 30	203
CCT Pro	GGG Gly	AAG Lys	ACT Thr	AAG Lys 35	ACT Thr	TTA Leu	GAT Asp	CCA Pro	AAG Lys 40	GCT Ala	GTC Val	TTT Phe	CAG Gln	AGA Arg 45	ACA Thr	251
			CTC Leu 50													266
(2)	INFO	ORMA1	NOI	FOR	SEQ	ID N	10: 1	109:								
	(i)	(<i>I</i> (E	A) LE B) T) C) S1	engti (PE : [Rani	HARACH: 31 NUCI DEDNE DGY:	l ba LEIC ESS:	ACII	pairs	5							
	(i.i	L) MC	DLEC	JLE 1	YPE:	CD	JA.									
	(v)	(7	A) OF	RGANI	SOURC ISM: E TYI	Homo	-		5							
	(i)	() ()	3) L(C) I(AME/E OCATI DENTI	KEY: ION: IFICA INFO	78. ATION	.170 N MET	THOD:	core							
	(x:	i) Si	EQUE	NCE I	DESC	RIPT	ON:	SEQ	ID 1	NO: :	109:					
ACT	GACG:	rgt (CTTT	GCTC	CT GA	ATACO	CATT	r TTC	CCCI	ACAC	CAC	ACCA	CTG 1	CCT	CTGTGC	60
CTG	rgga <i>i</i>	AAC (CACTO		ATG (Met 1					Pro :						110
CAT His -20	CTC Leu	ATC Ile	CGG Arg	TTG Leu	ATT Ile -15	ACT Thr	GTA Val	GCT Ala	TTC Phe	GGC Gly -10	CTG Leu	TAT Tyr	AAC Asn	CCC Pro	TCC Ser -5	158
TTA Leu	TGT Cys	CAT His	GCC Ala	TGT Cys 1	ACC Thr	AGA Arg	TGT Cys	TCC Ser 5	ACT Thr	GCA Ala	TCT Ser	GTA Val	TCC Ser 10	CAC His	CAG Gln	206
ATT Ile	GCA Ala	CAT His 15	TCC Ser	CCG Pro	AAG Lys	CAG Gln	AAA Lys 20	CCT Pro	TCT Ser	AAT Asn	CTG Leu	GGG Gly 25	GCC Ala	ATT Ile	CAG Gln	254

wo	WO 99/06439													PCT/IB9	8/01233
									85						
					CTA Leu										302
	ACA Thr														311
(2)	INFO	ORMAT	rion	FOR	SEQ	ID 1	NO: 1	110:							
(i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 458 base pairs (B) TYPE: NUCLEIC ACID															

(C) STRANDEDNESS: DOUBLE
(D) TOPOLOGY: LINEAR

(A) ORGANISM: Homo Sapiens(F) TISSUE TYPE: Pancreas

(A) NAME/KEY: sig_peptide
(B) LOCATION: 177..401

(D) OTHER INFORMATION: score 4.4

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 110:

(C) IDENTIFICATION METHOD: Von Heijne matrix

AATAATAGGC ACTGAAGACA TGTTAATGGA AGGTGGATTT GTGATTCAGA ACCTCTAGAC

GAA CCA ATA ACA TTC ACA GCA AGG AAA CAT CTG CTT CCT AAC GAG GTC Glu Pro Ile Thr Phe Thr Ala Arg Lys His Leu Leu Pro Asn Glu Val

TCG GTG GAT TTT GGC CTG CAG CTG GTG GGC TCC CTG CCT GTG CAT TCC

Ser Val Asp Phe Gly Leu Gln Leu Val Gly Ser Leu Pro Val His Ser

CTG ACC ACC ATG CCC ATG CTG CCC TGG GTT GTG GCT GAG GTG CGA AGA

Leu Thr Thr Met Pro Met Leu Pro Trp Val Val Ala Glu Val Arg Arg

CTC AGC AGG CAG TCC ACC AGA AAG GAA CCT GTA ACC ANG CAA NTC CGG

Leu Ser Arg Gln Ser Thr Arg Lys Glu Pro Val Thr Xaa Gln Xaa Arg

CTT TGC GTT TCA CCC TCT GGA CTG AGA TGT GAA CCT GAG CCA GGG AGA Leu Cys Val Ser Pro Ser Gly Leu Arg Cys Glu Pro Glu Pro Gly Arg

-35

-20 ·

TACCTGGGCG AGTCTTTTAA AATGTTTCTG CATATGAAGT GTGTAAAATA GATTGCTTGA 120

TCCAAAACAG AAAAACAGTG ATAACTGTTT TGCTGAGTTC CCCAAGACCCTT CCCAAG ATG 179

-65

seq QXRLCVSPSGLRC/EP

60

275

323

371

-75

-60

(ii) MOLECULE TYPE: CDNA

(vi) ORIGINAL SOURCE:

(ix) FEATURE:

-55

-40

-25

-10

	CAA CAG TGG GAT CCC CTG ATC TAT TCC AGC ATC TTT Gln Gln Trp Asp Pro Leu Ile Tyr Ser Ser Ile Phe 10 15	458
(2)	INFORMATION FOR SEQ ID NO: 111: (i) SEQUENCE CHARACTERISTICS:	
	(A) LENGTH: 177 base pairs (B) TYPE: NUCLEIC ACID (C) STRANDEDNESS: DOUBLE (D) TOPOLOGY: LINEAR	
	(ii) MOLECULE TYPE: CDNA	
	<pre>(vi) ORIGINAL SOURCE: (A) ORGANISM: Homo Sapiens (F) TISSUE TYPE: Colon</pre>	
	<pre>(ix) FEATURE: (A) NAME/KEY: sig_peptide (B) LOCATION: 58132 (C) IDENTIFICATION METHOD: Von Heijne matrix (D) OTHER INFORMATION: score 4.4</pre>	
	(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 111:	
	(iii) organist production. Organismo. III.	
AAA	AGGTACC CGCGAGAGAC AGCCAGCAGT TCTGTGGAGC AGCGGTGGCC GGCTAGG	57
ATG Met -25	GGC TGT CTC TGG GGT CTG GCT CTG CCC CTT TTC TTC	105
GAG Glu	GTT GGG GTC TCT GGG AGC TCT GCA GGC CCC AGC ACC CGC AGA GCA Val Gly Val Ser Gly Ser Ser Ala Gly Pro Ser Thr Arg Arg Ala -5 5	153
	ACT GCG ATG ACA ACG GAC GAC Thr Ala Met Thr Thr Asp Asp 10 15	177
(2)	INFORMATION FOR SEQ ID NO: 112:	
	(i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 175 base pairs (B) TYPE: NUCLEIC ACID (C) STRANDEDNESS: DOUBLE (D) TOPOLOGY: LINEAR	
	(ii) MOLECULE TYPE: CDNA	
	<pre>(vi) ORIGINAL SOURCE: (A) ORGANISM: Homo Sapiens (F) TISSUE TYPE: Lung (cells)</pre>	

(ix) FEATURE:

WO 99/06439 PCT/IB98/01233

87

 (A) NAME/KEY: sig_peptide (B) LOCATION: 62121 (C) IDENTIFICATION METHOD: Von Heijne matrix (D) OTHER INFORMATION: score 4.3 seq YLCHISLLDVTQQ/FP 	
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 112:	
ACAAGAAGAT CAGACACGTA CACTGAAAAA GAACCCACGT TACAGAGCAG TGTTGTTGAG	60
G ATG AAA CAA AAC ACA GAT CCA TAT CTG TGT CAT ATA AGC TTG CTC GAT Met Lys Gln Asn Thr Asp Pro Tyr Leu Cys His Ile Ser Leu Leu Asp -20 -15 -10 -5	109
GTA ACT CAA CAA TTT CCA AAT CCA CTT CCT GGC AGA ACC ATC TTT CCT Val Thr Gln Gln Phe Pro Asn Pro Leu Pro Gly Arg Thr Ile Phe Pro 1 5 10	157
GGT TCC TCA ACC CCC AGG Gly Ser Ser Thr Pro Arg 15	175
(2) INFORMATION FOR SEQ ID NO: 113: (i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 186 base pairs (B) TYPE: NUCLEIC ACID (C) STRANDEDNESS: DOUBLE (D) TOPOLOGY: LINEAR (ii) MOLECULE TYPE: CDNA (vi) ORIGINAL SOURCE: (A) ORGANISM: Homo Sapiens (D) DEVELOPMENTAL STAGE: Fetal (F) TISSUE TYPE: liver (ix) FEATURE: (A) NAME/KEY: sig_peptide (B) LOCATION: 34138 (C) IDENTIFICATION METHOD: Von Heijne matrix (D) OTHER INFORMATION: score 4.3 seq LISLLSSPNTPSA/SI	
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 113:	
ATAAATTTGT GTAAAATACA AAATATGTAA AAT ATG GTG ACA TAT TTT AAC TTC Met Val Thr Tyr Phe Asn Phe -35 -30	54
ACC TTC AAG CCA TTT TGC ATT CTG GCC TCA ATT ATT GTT CCC ACT CTT Thr Phe Lys Pro Phe Cys Ile Leu Ala Ser Ile Ile Val Pro Thr Leu -25	102
ATC TCT TTA CTT TCA TCT CCA AAT ACT CCA AGT GCA TCT ATT TAC TAT Ile Ser Leu Leu Ser Ser Pro Asn Thr Pro Ser Ala Ser Ile Tyr Tyr	150

PCT/IB98/01233

TCT CCA AAG TGT CTT TGT CCA TTA GCC ACC CCC AGG
Ser Pro Lys Cys Leu Cys Pro Leu Ala Thr Pro Arg
5 10 15

(2) INFORMATION FOR SEQ ID NO: 114:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 197 base pairs
 - (B) TYPE: NUCLEIC ACID
 - (C) STRANDEDNESS: DOUBLE
 - (D) TOPOLOGY: LINEAR
- (ii) MOLECULE TYPE: CDNA
- (vi) ORIGINAL SOURCE:
 - (A) ORGANISM: Homo Sapiens
 - (F) TISSUE TYPE: Colon
- (ix) FEATURE:

-10

- (A) NAME/KEY: sig_peptide
- (B) LOCATION: 114..182
- (C) IDENTIFICATION METHOD: Von Heijne matrix
- (D) OTHER INFORMATION: score 4.3

seq QXILLGTTSVVTA/AL

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 114:

(2) INFORMATION FOR SEQ ID NO: 115:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 377 base pairs
 - (B) TYPE: NUCLEIC ACID
 - (C) STRANDEDNESS: DOUBLE
 - (D) TOPOLOGY: LINEAR
- (ii) MOLECULE TYPE: CDNA
- (vi) ORIGINAL SOURCE:
 - (A) ORGANISM: Homo Sapiens
 - (F) TISSUE TYPE: Pancreas

(ix	 FEATURE:

- (A) NAME/KEY: sig_peptide
- (B) LOCATION: 105..233
- (C) IDENTIFICATION METHOD: Von Heijne matrix
- (D) OTHER INFORMATION: score 4.2

seq HMMAAAVADGTRA/AT

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 115:

AAG	CCCT	rgt '	rgaa(SACGO	CA GO	GCC	AACAG	GGG	SCCA!	ACGA	AGAT	rgac:	CT C	SATG	rcccgg	60
CCG1	GGT	ccc r	rctgi	rctg <i>i</i>	AG TA	ATGAT	rgcto	TAC	GAAA(GGA	GAA				A CAG A Gln -40	116
			AAA Lys													164
			GAG Glu -20													212
			GGG Gly													260
			TCT Ser													308
			CTG Leu													356
			CCC Pro													377

(2) INFORMATION FOR SEQ ID NO: 116:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 370 base pairs
 - (B) TYPE: NUCLEIC ACID
 - (C) STRANDEDNESS: DOUBLE
 - (D) TOPOLOGY: LINEAR
- (ii) MOLECULE TYPE: CDNA
- (vi) ORIGINAL SOURCE:
 - (A) ORGANISM: Homo Sapiens
 - (F) TISSUE TYPE: Lung (cells)
- (ix) FEATURE:
 - (A) NAME/KEY: sig_peptide
 - (B) LOCATION: 308..358
 - (C) IDENTIFICATION METHOD: Von Heijne matrix
 - (D) OTHER INFORMATION: score 4.2

seq SVIWFGSVXPCIS/XV

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 116:

TCTTTACAAT	TAGCTAGTTG	CGATTGGCTA	AGTTGTWAGG	GCCTAAATAC TTACTAAGCT	60
GTCTGGAGCC	TAACCCAGCA	GAAGGGTTAG	ATATTTCTTT	TAGCTTAGGG GCACAGTGAG	120
AAAAATTGCT	GAGACTAAGG	GCCCCATAAA	CATAGGAAAT	TGGAGAACTT CTAGCTGAAA	180
CTGATCATGG	TTCCTCCTTG	ATCAGTCTTG	GGGCAGGAAC	TGGGACTGGG GCCTGCCTCA	240
CCTGAAGCCC	ATGAAGAGTG	TATACCTGAA	TAAACTCTTA	TAAAGAAAGG AGGGGAATGC	300
				GGA TCT GTG WHC CCA Gly Ser Val Xaa Pro -5	349
	CAK GTT GA Xaa Val Gl				370

(2) INFORMATION FOR SEQ ID NO: 117:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 363 base pairs
 - (B) TYPE: NUCLEIC ACID
 - (C) STRANDEDNESS: DOUBLE
 - (D) TOPOLOGY: LINEAR
- (ii) MOLECULE TYPE: CDNA
- (vi) ORIGINAL SOURCE:
 - (A) ORGANISM: Homo Sapiens
 - (F) TISSUE TYPE: Lung
- (ix) FEATURE:
 - (A) NAME/KEY: sig peptide
 - (B) LOCATION: 253..342
 - (C) IDENTIFICATION METHOD: Von Heijne matrix
 - (D) OTHER INFORMATION: score 4.2

seq FLDFANLADLTLA/ES

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 117:

AAGAGGAAGG	AGCTGTGGGA AGCTCGCAGC AGGT	ATCGGA GCTTAAGCCA GTGGATTTGG 60
GGGCCCTGGG	CTCCCTAGCC GGCTGCGGTG TGAG	AATGGA GTGGGCAKGA AAGCAGCGGG 120
ACTTTCAGGT	AAGGGCAGCT CCGGGCTGGG ATCA	TTTGGC CTCCTTTCCT GGCCCTTCTC 180
TCCGGCTGTT	TTCTGGGAGT CARGMGAGTG TCTG	TAGTOT CTGCTCGGGG TTTGGGGCTC 240
AGGAATGATG		TCT ATT AGC TTA AGG AGG AGG 291 Ser Ile Ser Leu Arg Arg Arg -20

wo	WO 99/06439 91											РСТ/ІВ9	8/01233			
GAA Glu	ACA Thr	GCC Ala -15	AAT Asn	TTT Phe	CTT Leu	GAC Asp	TTT Phe -10	GCA Ala	AAT Asn	CTA Leu	GCT Ala	GAT Asp -5	CTC Leu	ACT Thr	CTT Leu	339
	GAA Glu 1															363
(2) INFORMATION FOR SEQ ID NO: 118: (i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 435 base pairs (B) TYPE: NUCLEIC ACID (C) STRANDEDNESS: DOUBLE (D) TOPOLOGY: LINEAR																
	(i:	L) MC	DLEC	JLE 1	YPE:	: CDI	NΑ									

(F) TISSUE TYPE: Lung (cells)

(vi) ORIGINAL SOURCE:

(ix) FEATURE: (A) NAME/KEY: other

(A) ORGANISM: Homo Sapiens

(B) LOCATION: 1..90

(C) IDENTIFICATION METHOD: blastn

(D) OTHER INFORMATION: identity 90 region 163..252 id AA236618

est

(ix) FEATURE:

(A) NAME/KEY: sig_peptide (B) LOCATION: 268..411

(C) IDENTIFICATION METHOD: Von Heijne matrix

(D) OTHER INFORMATION: score 4.1

seq FTTLSNLSLPSQT/KD

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 118:

CGTCCATCAG ATCCTGACCC TAAGGCTGGT GACTTTGTAG CTATGAAAGC AGGAAATGAA	60
GGCATAGTAC AATTTCCTAA AGATGAAAAA CAATATCTTG TTCCCCTCTG TTTTTGTTAT	120
TACAAGAAAT AATCTATCTA CTACTCGTCA GCACCTGTCT TTGTGTCTGA AGCCAAGAAT	180
AAATTCATCA ACTTGGTAGC CACTGCTGCA ACTGAAGCCA ACCGCAGTCA ATGTTGGCTA	240
TATGTTGAGT TGCCGGAGGC GCCTGGA ATG TGC TAC CTT GCA GAA TTG TCC CTG Met Cys Tyr Leu Ala Glu Leu Ser Leu -45 -40	294
ACA ACA TTT CKV MAT GGC TAT ATT GTT ACC AGT AGG GCC ACA ACA ACA Thr Thr Phe Xaa Xaa Gly Tyr Ile Val Thr Ser Arg Ala Thr Thr Thr -35 -30 -25	342
ACA ACA CTT GCA ATC CAA CCT GGG CTT CCT TTC ACC ACA CTA AGC AAT Thr Thr Leu Ala Ile Gln Pro Gly Leu Pro Phe Thr Thr Leu Ser Asn	390

92

-20 -15 -10

CTA TCT TTG CCA AGT CAG ACA AAA GAT GAA CTC CAC CCT CCC TGG
Leu Ser Leu Pro Ser Gln Thr Lys Asp Glu Leu His Pro Pro Trp
-5 1 5

(2) INFORMATION FOR SEQ ID NO: 119:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 127 base pairs
- (B) TYPE: NUCLEIC ACID
- (C) STRANDEDNESS: DOUBLE
- (D) TOPOLOGY: LINEAR
- (ii) MOLECULE TYPE: CDNA
- (vi) ORIGINAL SOURCE:
 - (A) ORGANISM: Homo Sapiens
 - (F) TISSUE TYPE: Pancreas
- (ix) FEATURE:
 - (A) NAME/KEY: sig_peptide
 - (B) LOCATION: 56..106
 - (C) IDENTIFICATION METHOD: Von Heijne matrix
 - (D) OTHER INFORMATION: score 4.1

seq LSSLILLPIWINM/AQ

- (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 119:
- ATGCCCCCTC AGCCCTGTGG CTGGGGGCAG AGCTCAGACT GTCTTCTGAA GATTG ATG 58

TCT ATT TCC TTG AGC TCT TTA ATT TTG TTG CCA ATT TGG ATA AAC ATG

Ser Ile Ser Leu Ser Ser Leu Ile Leu Leu Pro Ile Trp Ile Asn Met

-15 **-**10 -5

GCA CAA ATC CAG CGG GGA GGT
Ala Gln Ile Gln Arg Gly Gly
1 5

- (2) INFORMATION FOR SEQ ID NO: 120:
 - (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 462 base pairs
 - (B) TYPE: NUCLEIC ACID
 - (C) STRANDEDNESS: DOUBLE
 - (D) TOPOLOGY: LINEAR
 - (ii) MOLECULE TYPE: CDNA
 - (vi) ORIGINAL SOURCE:
 - (A) ORGANISM: Homo Sapiens
 - (F) TISSUE TYPE: Lung
 - (ix) FEATURE:
 - (A) NAME/KEY: sig_peptide

(B) LOCATION: 88..345

(D) TOPOLOGY: LINEAR

(A) ORGANISM: Homo Sapiens(F) TISSUE TYPE: Large intestine

(ii) MOLECULE TYPE: CDNA

(vi) ORIGINAL SOURCE:

(D) OTHER INFORMATION: score 4

93

(C) IDENTIFICATION METHOD: Von Heijne matrix

	seq LLFALSWKSDAPA/TS											·				
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 120:																
ACG	CTTC	CTG :	racc <i>i</i>	ACCCI	rg ci	CAAC	STAGO	GG#	ACAC	GGAA	CAG	GAAG	CTA 1	rcago	CCCGTC	60
GGCCTCCGGG CCCTGCATTC TCTAGCC ATG GAC CGG GAC CTT TTG CGG CAG TCG 11 Met Asp Arg Asp Leu Leu Arg Gln Ser -85 -80												114				
CTA Leu	AAT Asn	TGC Cys -75	CAC His	GGG Gly	TCG Ser	TCT Ser	TTG Leu -70	CTC Leu	TCT Ser	CTA Leu	CTT Leu	CGG Arg -65	AGC Ser	GAA Glu	CAG Gln	162
CAG Gln	GAC Asp -60	AAT Asn	CCA Pro	CAC His	TTC Phe	CGT Arg -55	AGC Ser	CTC Leu	CTG Leu	GGG Gly	TCG Ser -50	GCC Ala	GCC Ala	GAG Glu	CCA Pro	210
GCC Ala -45	CGG Arg	GGC Gly	CCG Pro	CCG Pro	CCC Pro -40	CAG Gln	CAC His	CCG Pro	TTG Leu	CAG Gln -35	GGC Gly	AGA Arg	AAA Lys	GAG Glu	AAG Lys -30	258
AGA Arg	GTT Val	GAC Asp	AAC Asn	ATC Ile -25	GAG Glu	ATA Ile	CAG Gln	AAA Lys	TTC Phe -20	ATC Ile	TCC Ser	AAA Lys	AÄA Lys	GCG Ala -15	GAT Asp	306
	CTT Leu															354
	AAT Asn 5															402
CAA Gln 20	TTC Phe	ATG Met	GAG Glu	ATA Ile	CCT Pro 25	AGT Ser	ATG Met	GAT Asp	CGG Arg	AGA Arg 30	GAG Glu	CTG Leu	TTT Phe	TTC Phe	CGA Arg 35	450
	ATT Ile															462
(2)	INF	ORMA'	TION	FOR	SEQ	ID I	NO:	121:						•		
	(i	(,	A) L B) T	ENGT:	HARACH: 3: NUC: DEDNI	20 b	ase ACI	pair. D	s							

PCT/IB98/01233

94

WO 99/06439

(ix) FEATURE:

(A) NAME/KEY: sig_peptide (B) LOCATION: 246296 (C) IDENTIFICATION METHOD: Von Heijne matrix (D) OTHER INFORMATION: score 4 seq TLVTXXNASCSFA/SV (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 121:	
GGTTTCTTGG TTTTCTTGTT TTCCCATATG TTTGTATCCC TAAACAAAAG ATTGCTTAGT	60
CTGCATGTT TTTCAACTTT ATGTAAATGA AATCACACTG CATTTATTCT TCTCTGACTT	120
PATTTAGCTG AACATTATGC ATCTGATCCC CATCCATATT GTCATGGGTA ACTAGCATTT	180
ATTGTCTTCA CTGCTGAACA AGTAGAACCT ATTCAGTTCA CTGCCCACTT GTCTTCCCAG	240
ATTCC ATG GTT CTG GCC ACA CTA GTG ACT TST KTA AAT GCA TCC TGC TCT Met Val Leu Ala Thr Leu Val Thr Xaa Xaa Asn Ala Ser Cys Ser -15 -10 -5	290
TTT GCG TCT GTG CAT CTT GCC CAG GGT GGG Phe Ala Ser Val His Leu Ala Gln Gly Gly 1 5	320
(2) INFORMATION FOR SEQ ID NO: 122: (i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 283 base pairs (B) TYPE: NUCLEIC ACID (C) STRANDEDNESS: DOUBLE (D) TOPOLOGY: LINEAR (ii) MOLECULE TYPE: CDNA (vi) ORIGINAL SOURCE: (A) ORGANISM: Homo Sapiens (F) TISSUE TYPE: Lung (cells) (ix) FEATURE: (A) NAME/KEY: sig_peptide (B) LOCATION: 167226 (C) IDENTIFICATION METHOD: Von Heijne matrix (D) OTHER INFORMATION: score 4 seq IILKVLLNQTCQT/VQ (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 122:	
AAAAGGTCTT CGAATTTCAA CTCGTCAAGT ACATCACACT CCAGTTTGAA AAGTCCAAGC	60
CACATGGAAA AATACCCGCA AAAAGAGAAA ACCAAAGAAG ATCTGGATTC ACGAAGCAAC	120
CTACACTTGC CAGAAACTAA ATTTTCTGAA TTGTCAAAAC TGAAGA ATG ATG ATA Met Met Ile -20	175
TGG AAA AGG CTA ATC ATA TTG AAA GTG TTA TTA AAT CAA ACT TGC CAA	223

95

Trp Lys Arg Leu IIe II -15	e Leu Lys Val 1 -10	Leu Leu Asn Gln 1 -5	Thr Cys Gln
ACT GTG CAA ACA GTG AC Thr Val Gln Thr Val Th 1	A CCG ACT TCA :	TGG GTC TTT TCA P Trp Val Phe Ser P 10	AAT CAA GCC 271 Asn Gln Ala 15
GGT ATG ACC CGG Gly Met Thr Arg			283
(2) INFORMATION FOR SE	O ID NO: 123:		
(i) SEQUENCE CHAR (A) LENGTH: (B) TYPE: NU (C) STRANDED (D) TOPOLOGY	322 base pairs CLEIC ACID NESS: DOUBLE	·	
(ii) MOLECULE TYP	E: CDNA		
	: Homo Sapiens ENTAL STAGE: Fe	etal	
(B) LOCATION	CATION METHOD: FORMATION: sc	Von Heijne matr: ore 4 q AALVKCLPVLCLA/0	
(xi) SEQUENCE DES	CRIPTION: SEQ	ID NO: 123:	
GCTCTGGGGC TATCAGGCCA		GGC AAA GCG GGG Gly Lys Ala Gly -55	
AAG ACT CAC TGC TCA GC Lys Thr His Cys Ser Al -50	a Gln Arg Pro	GAT GTC TGC AGG ' Asp Val Cys Arg ' -40	TGG CTG AGC 100 Trp Leu Ser
CCC TTC ATC CTC TCC TG Pro Phe Ile Leu Ser Cy -35 -3	s Cys Val Tyr	TTC TGC CTC TGG A Phe Cys Leu Trp	ATT CCC GAG 148 Ile Pro Glu -20
GAC CAG CTG TCC TGG TT Asp Gln Leu Ser Trp Ph -15	e Ala Ala Leu	GTC AAG TGC CTG (Val Lys Cys Leu -10	CCC GTC CTC 196 Pro Val Leu -5
TGC CTG GCT GGG TTC CT Cys Leu Ala Gly Phe Le 1	G TGG GTC ATG u Trp Val Met 5	TCC CCA AGC GGG (Ser Pro Ser Gly (GGC TAC ACC 244 Gly Tyr Thr
CAG CTC CTC CAG GGA GC Gin Lau Leu Gln Gly Al	C CTT GTG TGC a Leu Val Cys	TCG GCT GTG GGG Ser Ala Val Gly	GAC GCT TGC 292 Asp Ala Cys

WO 99/06439	PCT/IB98/012
96	
CTC ATC TGG CCG GCA GCC TTC GTC CCA GGG Leu Ile Trp Pro Ala Ala Phe Val Pro Gly 30 35	322
(2) INFORMATION FOR SEQ ID NO: 124:	
(i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 316 base pairs (B) TYPE: NUCLEIC ACID (C) STRANDEDNESS: DOUBLE (D) TOPOLOGY: LINEAR	
(ii) MOLECULE TYPE: CDNA	
<pre>(vi) ORIGINAL SOURCE: (A) ORGANISM: Homo Sapiens (F) TISSUE TYPE: Lung (cells)</pre>	
<pre>(ix) FEATURE: (A) NAME/KEY: sig_peptide (B) LOCATION: 107190 (C) IDENTIFICATION METHOD: Von Heijne matrix</pre>	

seq PLLGVLFFQGVYI/VF

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 124:

(D) OTHER INFORMATION: score 4

ATT	TAAGT	TG 1	rcrc:	CGG	CC TO	CCTG	CTCTC	ACC	STCAC	CTTC	CGGT	rgtt <i>i</i>	ACC 1	rgtgi	CGTTA	60
CCG	GAGO	CTG :)AAA1	CAAGO	ST G	rgca <i>i</i>	AGCAT	CTO	SAAGA	AGCT	GCC			CAG (115
														CTG Leu		163
														ATT Ile		211
GCA Ala	GAT Asp	GCC Ala 10	CAT His	GTC Val	CGA Arg	GGT Gly	TAT Tyr 15	GTT Val	GGA Gly	GAA Glu	AAG Lys	ATC Ile 20	AAG Lys	TTG Leu	AAA Lys	259
														ATA Ile		307
	ACA Thr										•					316

- (2) INFORMATION FOR SEQ ID NO: 125:
 - (i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 153 base pairs

(B)	TYPE:	NUCLEIC	ACID	

- (C) STRANDEDNESS: DOUBLE
- (D) TOPOLOGY: LINEAR
- (ii) MOLECULE TYPE: CDNA
- (vi) ORIGINAL SOURCE:
 - (A) ORGANISM: Homo Sapiens
 - (F) TISSUE TYPE: Colon
- (ix) FEATURE:
 - (A) NAME/KEY: sig_peptide
 - (B) LOCATION: $1..\overline{9}6$
 - (C) IDENTIFICATION METHOD: Von Heijne matrix
 - (D) OTHER INFORMATION: score 4

seq LIYWYVLLILSFP/FI

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 125:

ATG CTC GGA ACA CAT ATC TAC GTG TCC TTA TGG ATT ATT CTT TTC TCT

Met Leu Gly Thr His Ile Tyr Val Ser Leu Trp Ile Ile Leu Phe Ser

-30

-25

-20

TCC CCA CAT CTT ATC TAT TGG TAT GTT CTG TTG ATT TTG TCT TTT CCA

Ser Pro His Leu Ile Tyr Trp Tyr Val Leu Leu Ile Leu Ser Phe Pro

-15

-10

-5

TTT ATC ATC AAA TTT TCT ATG AAC ACC TTG TCC AGA CCA CCT GAC

Phe Ile Ile Lys Phe Ser Met Asn Thr Leu Ser Arg Pro Pro Asp

1 5 10 15

ACC CCC CAG
Thr Pro Gln
153

- (2) INFORMATION FOR SEQ ID NO: 126:
 - (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 131 base pairs
 - (B) TYPE: NUCLEIC ACID
 - (C) STRANDEDNESS: DOUBLE
 - (D) TOPOLOGY: LINEAR
 - (ii) MOLECULE TYPE: CDNA
 - (vi) ORIGINAL SOURCE:
 - (A) ORGANISM: Homo Sapiens
 - (F) TISSUE TYPE: Pancreas
 - (ix) FEATURE:
 - (A) NAME/KEY: sig_peptide
 - (B) LOCATION: 69..125
 - (C) IDENTIFICATION METHOD: Von Heijne matrix
 - (D) OTHER INFORMATION: score 3.9

seq FLNLHGFLGHLLS/GE

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 126:

ACATCAGGAT AAAAAATCTG TAATACTAAA AATGTTAAAT AATTTCAGTT GCCAAATTTT	60
CAGTTGAA ATG TCA ATA TAT AAT TTA TTT CTT AAT TTG CAT GGC TTT TTA Met Ser Ile Tyr Asn Leu Phe Leu Asn Leu His Gly Phe Leu -15 -10	110
GGT CAT TTA TTA TCT GGG GAG Gly His Leu Leu Ser Gly Glu -5 1	131
(2) INFORMATION FOR SEQ ID NO: 127:	
(i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 429 base pairs (B) TYPE: NUCLEIC ACID (C) STRANDEDNESS: DOUBLE (D) TOPOLOGY: LINEAR	
(ii) MOLECULE TYPE: CDNA	
<pre>(vi) ORIGINAL SOURCE: (A) ORGANISM: Homo Sapiens (F) TISSUE TYPE: Lung (cells)</pre>	
<pre>(ix) FEATURE: (A) NAME/KEY: sig_peptide (B) LOCATION: 343417 (C) IDENTIFICATION METHOD: Von Heijne matrix (D) OTHER INFORMATION: score 3.9</pre>	
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 127:	
CCTTGCTTCC STAACAGACT TTAGCCCTAG GGAAGTCAGA CCTGATCAGT GCAGGGCGTT	60
CTTGCCGATA AGTTGGGGAT GGTTGGACCT GAGCACCTCT TGGTCAGCTG GCCTCTCCCA	120
GGGCTCCAGG CTACCCACAC CTGCTTGCAG TGCGACCTCG GTTGCCCAGG TGGGGTCCCT	180
TATGGGGGTG TACATCATAG CTCCTGTGCT GGTGAACTGC AGTTGATAAA TCCCATAGTC	240
GTGCAACAGC CCACCTGCAC CTTCCCAAAA CTGCGGCCTC TCTGAGACAC TTTGCCTGCA	300
TGCATTCACC TATGGCCACC CCCGACGTTG CTTTGCCAGC AT ATG TGC ATG CAA Met Cys Met Gln -25	354
GTG GAC CTT GCC TTC TCT TTT CCA CCA GCA TGC GTG TGC ATG TGC ACC Val Asp Leu Ala Phe Ser Phe Pro Pro Ala Cys Val Cys Met Cys Thr -20 -15 -10	402
CKG TCA TGC TAC AGC TGC CAG TGT GAG Xaa Ser Cys Tyr Ser Cys Gln Cys Glu -5	429

99

(i)	SEQUI	ENCE CHAI	RACTI	ERIST	cs:
	(A)	LENGTH:	265	base	pairs

(B) TYPE: NUCLEIC ACID
(C) STRANDEDNESS: DOUBLE

(D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: CDNA

(vi) ORIGINAL SOURCE:

(A) ORGANISM: Homo Sapiens(F) TISSUE TYPE: Pancreas

• •

(ix) FEATURE:

(A) NAME/KEY: sig_peptide

(B) LOCATION: 56..151

(C) IDENTIFICATION METHOD: Von Heijne matrix

(D) OTHER INFORMATION: score 3.8

seq IRTATLVISLARG/WQ

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 128:

AACA	ACTC	ATT I	TAGO	CAG	rg to	CCAGO	CTAI	CAC	CAGA	AGAA	AGA	CAGG1	rgg (GCAGO	ATG Met	58
						AGC Ser -25										106
						ACC Thr										154
						AAC Asn										202
						ACC Thr										250
	ACA Thr															265

(2) INFORMATION FOR SEQ ID NO: 129:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 222 base pairs
 - (B) TYPE: NUCLEIC ACID
 - (C) STRANDEDNESS: DOUBLE
 - (D) TOPOLOGY: LINEAR
- (ii) MOLECULE TYPE: CDNA
- (vi) ORIGINAL SOURCE:
 - (A) ORGANISM: Homo Sapiens
 - (D) DEVELOPMENTAL STAGE: Fetal

			100
			100

(F) TISSUE TYPE: liver	
<pre>(ix) FEATURE: (A) NAME/KEY: sig_peptide</pre>	
(B) LOCATION: 67192(C) IDENTIFICATION METHOD: Von Heijne matrix	
(D) OTHER INFORMATION: score 3.7	
seq PLLHLFYQHLCFP/VP	
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 129:	
ACAACTTTAT GTCTCAAATG TTTGTATTAG AGATTTTTCT TGAAGTAAAA CCTACCTCTA	60
CTAGGA ATG GAA CCA AAA AGG GGG AGR ATG TGG TKA TTT GAA ATT GAA Met Glu Pro Lys Arg Gly Arg Met Trp Xaa Phe Glu Ile Glu -40 -35 -30	108
GAT AGC TGT ATA TAC CAG GAC ATC CCA TCG TTT GTC TTA CTT TAC CCA Asp Ser Cys Ile Tyr Gln Asp Ile Pro Ser Phe Val Leu Leu Tyr Pro -25 -20 -15	156
CTT CTC CAT TTG TTT TAC CAG CAT CTC TGT TTT CCT GTT CCA TGC ACT Leu Leu His Leu Phe Tyr Gln His Leu Cys Phe Pro Val Pro Cys Thr -10 -5 1	204
AGA AAT CCT GGG CCC GGG Arg Asn Pro Gly Pro Gly	222
5 10	
(2) INFORMATION FOR SEQ ID NO: 130:	
(i) SEQUENCE CHARACTERISTICS:	
(A) LENGTH: 211 base pairs (B) TYPE: NUCLEIC ACID	
(C) STRANDEDNESS: DOUBLE	
(D) TOPOLOGY: LINEAR	
(ii) MOLECULE TYPE: CDNA	
(vi) ORIGINAL SOURCE:	
(A) ORGANISM: Homo Sapiens (F) TISSUE TYPE: Lung (cells)	
<pre>(ix) FEATURE: (A) NAME/KEY: sig_peptide</pre>	
(B) LOCATION: 131181(C) IDENTIFICATION METHOD: Von Heijne matrix	
(D) OTHER INFORMATION: score 3.6	
seq SVLQRCLFSFVTS/VF	
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 130:	
CACATTCTGC ACTCTAGCCC AGAAGTCACA AACTGATGGC TCTAGTGTCA ATTCTGCCTG	60
TAGTGCTGAC CACTCATACA GCTGGCCTGC TTCCCTGATT TATCTATTAG CAACTCCTAC	120
TTCTGGCCAT ATG GAA TTT TGT TCA GTT CTT CAA AGG TGC CTA TTC TCC	169

110 33100 103		U. U.
•	101	
Ме	et Glu Phe Cys Ser Val Leu Gln Arg Cys Leu Phe Ser -15 -10 -5	
TTT GTC ACT The Val Thr S	TCG GTC TTT CAT ATG CTG TTC CCT CTG CCT GGG Ser Val Phe His Met Leu Phe Pro Leu Pro Gly 1 5 10	211
(2) INFORMAT	ION FOR SEQ ID NO: 131:	•
(A) (B) (C)	UENCE CHARACTERISTICS:) LENGTH: 264 base pairs) TYPE: NUCLEIC ACID) STRANDEDNESS: DOUBLE) TOPOLOGY: LINEAR	
(ii) MOI	LECULE TYPE: CDNA	
(A) (D)	IGINAL SOURCE:) ORGANISM: Homo Sapiens) DEVELOPMENTAL STAGE: Fetal) TISSUE TYPE: liver	
(B)	ATURE:) NAME/KEY: sig_peptide) LOCATION: 196249) IDENTIFICATION METHOD: Von Heijne matrix) OTHER INFORMATION: score 3.6 seq IYVLLFFLLMKFS/FD	
(xi) SE(QUENCE DESCRIPTION: SEQ ID NO: 131:	
TAAACTAAGA G	TGACATGAC CAGATTTGGA CTTGACAAAG GTGCCTAGAA ATGAATGAAG	60
GACAGTTATA AC	CAGTTCACC CACAGAAACA TAGAGTGCAT TTCTGTCGAT ATCAAAGGAA	120
CTTAGTGGAG A	AATGACAAC GACCTGGATG AAAGCCCTGG CTTGACTTGG GGACCCAGTA	180
CAGGGTCAAC A	AACT ATG GCA GAG AGC CAA ATC TAC GTG CTG CTT TTT TTT Met Ala Glu Ser Gln Ile Tyr Val Leu Phe Phe -15 -10	23:
	AAG TTT TCT TTT GAC ACA CGA GGG Lys Phe Ser Phe Asp Thr Arg Gly 1 5	264
(2) INFORMAT	ION FOR SEQ ID NO: 132:	
(A (B (C	UENCE CHARACTERISTICS:) LENGTH: 93 base pairs) TYPE: NUCLEIC ACID) STRANDEDNESS: DOUBLE) TOPOLOGY: LINEAR	

(ii) MOLECULE TYPE: CDNA

(vi) ORIGINAL SOURCE:

102

(A) ORGANISM: Homo Sapiens (F) TISSUE TYPE: Liver

() X	FEATURE	

- (A) NAME/KEY: sig_peptide
- (B) LOCATION: 13..66
- (C) IDENTIFICATION METHOD: Von Heijne matrix
- (D) OTHER INFORMATION: score 3.6

seq ACSLSSGPLQINA/LP

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 132:

AAAAAACTCT GA ATG CAG ACA AAC AAT GCT TGT TCC CTG TCC TCT GGC CCT 51 Met Gln Thr Asn Asn Ala Cys Ser Leu Ser Ser Gly Pro -15 -10

TTG CAA ATA AAT GCC TTA CCA GAC CTG CCC TGC CAC CCC GGG 93 Leu Gln Ile Asn Ala Leu Pro Asp Leu Pro Cys His Pro Gly

(2) INFORMATION FOR SEQ ID NO: 133:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 390 base pairs
 - (B) TYPE: NUCLEIC ACID
 - (C) STRANDEDNESS: DOUBLE
 - (D) TOPOLOGY: LINEAR
- (ii) MOLECULE TYPE: CDNA
- (vi) ORIGINAL SOURCE:
 - (A) ORGANISM: Homo Sapiens
 - (F) TISSUE TYPE: Large intestine
- (ix) FEATURE:
 - (A) NAME/KEY: sig_peptide
 - (B) LOCATION: 304..381
 - (C) IDENTIFICATION METHOD: Von Heijne matrix
 - (D) OTHER INFORMATION: score 3.6

seq KVLMGLLCNQTAA/KR

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 133:

CAAATATTCC CTCCTCTGGG GAACTTCTAA TTCCCCAGCA GAAATCATGC CTCCAGTGAC GGTCCATGCA GGCTTCTGGA GCGTTGGGCA ATCTTCCTCA TCTGTGGACT TGCTGGCCTC 120 TTTCCCCTTG GCTTCTTGAG GCCTGCGCTG CATCTGATTT CTCTGTCTCG GGGGCCTAGC 180 ATGGTCATTG GCCCTCCCAG TGTTTCCTGG ATGATCATCG TGCTGTTCCT GAGTCAGGGC 240 TGCCATTGGA GGTGACATCT GTGACTGCAA CCTGTGCCTG AATTGGTGGG CGGAACCTGC 300 TCA ATG GGC CAG AAC AAT GCT TCC TTC CAC TGC CCC TGC CTG AAA GTC

Met Gly Gln Asn Asn Ala Ser Phe His Cys Pro Cys Leu Lys Val

WO 99/06439	PCT/IB98/01233

CTC ATG GGC CTC CTT TGC AAT CAA ACT GCT GCC AAG AGA CCT Leu Met Gly Leu Leu Cys Asn Gln Thr Ala Ala Lys Arg Pro -10 -5 1	390
(2) INFORMATION FOR SEQ ID NO: 134:	
(i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 124 base pairs (B) TYPE: NUCLEIC ACID (C) STRANDEDNESS: DOUBLE (D) TOPOLOGY: LINEAR	
(ii) MOLECULE TYPE: CDNA	
<pre>(vi) ORIGINAL SOURCE: (A) ORGANISM: Homo Sapiens (F) TISSUE TYPE: Liver</pre>	
<pre>(ix) FEATURE: (A) NAME/KEY: sig_peptide (B) LOCATION: 2970 (C) IDENTIFICATION METHOD: Von Heijne matrix (D) OTHER INFORMATION: score 3.6</pre>	
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 134:	
ACACCTTTCT ATCTGACTAC CTGCCCTA ATG TTG CCC TTA CTC TCA GTA ATG Met Leu Pro Leu Leu Ser Val Met -10	52
TGG TCT CCT ATT GCG CCA CTC ACA GTA GGA TCC AAG GAC CCA TGC CAC Trp Ser Pro Ile Ala Pro Leu Thr Val Gly Ser Lys Asp Pro Cys His -5 1 5 10	100
TTT ATA CCA GTT CAT GAC GAG ATG Phe Ile Pro Val His Asp Glu Met 15	124
(2) INFORMATION FOR SEQ ID NO: 135:	
(i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 179 base pairs (B) TYPE: NUCLEIC ACID (C) STRANDEDNESS: DOUBLE (D) TOPOLOGY: LINEAR	
(ii) MOLECULE TYPE: CDNA	
<pre>(vi) ORIGINAL SOURCE: (A) ORGANISM: Homo Sapiens (F) TISSUE TYPE: Thyroid</pre>	
<pre>(ix) FEATURE: (A) NAME/KEY: sig_peptide (B) LOCATION: 87155</pre>	

104

(C) IDENTIFICATION METHOD: Von Heijne matrix(D) OTHER INFORMATION: score 3.6seq WITCPPTFHGCRA/LF	
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 135:	
ATTCAAAGTA CATTTGACAA CCCACTGCAA GTTGTGGCAT ACATGGGTGC CATGAACCAT	60
GATACCAACT ACAGCTTTCA GGTTCA ATG TGG CTT AAT TGT GGT GGC CTA CAA Met Trp Leu Asn Cys Gly Gly Leu Gln -20 -15	113
AGA TGG ATC ACC TGC CCA CCC ACA TTT CAT GGA TGC AGA GCT CTG TTC Arg Trp Ile Thr Cys Pro Pro Thr Phe His Gly Cys Arg Ala Leu Phe -10 -5 1	161
CCA GTA CTG GAC GCC GGG Pro Val Leu Asp Ala Gly 5	179
(2) INFORMATION FOR SEQ ID NO: 136: (i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 145 base pairs (B) TYPE: NUCLEIC ACID (C) STRANDEDNESS: DOUBLE (D) TOPOLOGY: LINEAR (ii) MOLECULE TYPE: CDNA (vi) ORIGINAL SOURCE:	
(A) ORGANISM: Homo Sapiens (F) TISSUE TYPE: Lung (cells) (ix) FEATURE: (A) NAME/KEY: sig_peptide (B) LOCATION: 26121 (C) IDENTIFICATION METHOD: Von Heijne matrix (D) OTHER INFORMATION: score 3.6 seq LGVLTFILQRTTC/LN	
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 136:	
ACTCTGCCGG GTGGATGCTT TCTCC ATG TGG CAA GGC TGT AAC TGT TCA CAG Met Trp Gln Gly Cys Asn Cys Ser Gln -30 -25	52
CTG TCT GAA ACA GCA GTG GAC CAG GAG CAG CTT GGA GTT TTA ACT TTC Leu Ser Glu Thr Ala Val Asp Gln Glu Gln Leu Gly Val Leu Thr Phe -20 -15 -10	100
ATT TTA CAA AGA ACA ACA TGT TTG AAT GTT TCA GCA GGC AAG AGG Ile Leu Gln Arg Thr Thr Cys Leu Asn Val Ser Ala Gly Lys Arg -5 1 5	145

WO 99/06439 105

	•	
(2)	INFORMATION FOR SEQ ID NO: 137:	
	(i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 211 base pairs (B) TYPE: NUCLEIC ACID (C) STRANDEDNESS: DOUBLE (D) TOPOLOGY: LINEAR	
	(ii) MOLECULE TYPE: CDNA	
	<pre>(vi) ORIGINAL SOURCE: (A) ORGANISM: Homo Sapiens (F) TISSUE TYPE: Lung (cells)</pre>	
	(ix) FEATURE: (A) NAME/KEY: sig_peptide (B) LOCATION: 86196 (C) IDENTIFICATION METHOD: Von Heijne matrix (D) OTHER INFORMATION: score 3.6 seq LVTLLASKSPSCP/LH	
	(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 137:	
AGGG	GCTGTK DGTTGCTGCA GCAACACTGA CKDATACACT GCAAGCTCAG TGTATTAGTC	60
AGCT	CTGATA TTCTAGCATT CTTGA ATG TGT CTC CCC CAT CCC CAG GTT GTC Met Cys Leu Pro His Pro Gln Val Val -35 -30	112
TCT Ser	TCT AAT TTT CAC ATC CTC ATC TTC CTC CTG CCC ACG AAG ATG CTC Ser Asn Phe His Ile Leu Ile Phe Leu Leu Pro Thr Lys Met Leu -25 -2015	160
GTG Val	ACA CTT CTC GCH TCA AAA TCT CCG AGT TGC CCC CTT CAC CCC CTA Thr Leu Leu Ala Ser Lys Ser Pro Ser Cys Pro Leu His Pro Leu -10 -5 1	208
CGG Arg 5		211
(2)	INFORMATION FOR SEQ ID NO: 138:	
	(i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 333 base pairs (B) TYPE: NUCLEIC ACID (C) STRANDEDNESS: DOUBLE (D) TOPOLOGY: LINEAR	
	(ii) MOLECULE TYPE: CDNA	
	<pre>(vi) ORIGINAL SOURCE: (A) ORGANISM: Homo Sapiens (F) TISSUE TYPE: Pancreas</pre>	
	(ix) FEATURE: (A) NAME/KEY: sig_peptide (B) LOCATION: 1222	

106

(C) IDENTIFICATION METHOD: Von Heijne matrix
(D) OTHER INFORMATION: score 3.5
seq ECLNLLLSSGADL/RR

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 138:

		CTC Leu						48
		TTG Leu						96
		GGG Gly						144
		GCT Ala						192
		AGT Ser -5						240
		SAC Xaa						288
		ACT Thr						333

(2) INFORMATION FOR SEQ ID NO: 139:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 107 base pairs
 - (B) TYPE: NUCLEIC ACID
 - (C) STRANDEDNESS: DOUBLE
 - (D) TOPOLOGY: LINEAR
- (ii) MOLECULE TYPE: CDNA
- (vi) ORIGINAL SOURCE:
 - (A) ORGANISM: Homo Sapiens
 - (F) TISSUE TYPE: Thyroid
- (ix) FEATURE:
 - (A) NAME/KEY: sig_peptide
 - (B) LOCATION: 9..80
 - (C) IDENTIFICATION METHOD: Von Heijne matrix
 - (D) OTHER INFORMATION: score 3.5 seq LHDCFLSVFQVLS/SI
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 139:

									107					
GAAG	SATGG						Cys				Gl		r GAC s Asp	50
						CAA Gln								98
	CTT Leu													107
(2)	INFO	RMAT	ON	FOR	SEQ	ID N	0: 1	40:						

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 220 base pairs
 - (B) TYPE: NUCLEIC ACID
 - (C) STRANDEDNESS: DOUBLE
 - (D) TOPOLOGY: LINEAR
- (ii) MOLECULE TYPE: CDNA
- (vi) ORIGINAL SOURCE:
 - (A) ORGANISM: Homo Sapiens
 - (F) TISSUE TYPE: Thyroid
- (ix) FEATURE:
 - (A) NAME/KEY: sig_peptide
 - (B) LOCATION: 50..109
 - (C) IDENTIFICATION METHOD: Von Heijne matrix
 - (D) OTHER INFORMATION: score 3.5

seq KFCLICLLTF1FH/HC

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 140:

GAGCGGAGAG GCAGATGCAC ACGGCACTCG AGTGTGAGGA AAAATAGAA ATG AAG GTA Met Lys Val -20																
					TTT Phe											106
CAT His	CAT His 1	TGC Cys	AAC Asn	CAT His	TGC Cys 5	CAT His	GAA Glu	GAA Glu	CAT His	GAC Asp 10	CAT His	GGC Gly	CCT Pro	GAA Glu	GCG Ala 15	154
					CGT Arg											202
			GCC Ala 35													220

(vi) ORIGINAL SOURCE:

(i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 317 base pairs (B) TYPE: NUCLEIC ACID (C) STRANDEDNESS: DOUBLE (D) TOPOLOGY: LINEAR (ii) MOLECULE TYPE: CDNA														
	(ii) M	DLECUL	E TYPE:	CDNA	Ą									
	(vi) ORIGINAL SOURCE: (A) ORGANISM: Homo Sapiens (F) TISSUE TYPE: Lung													
<pre>(ix) FEATURE: (A) NAME/KEY: sig_peptide (B) LOCATION: 87161 (C) IDENTIFICATION METHOD: Von Heijne matrix (D) OTHER INFORMATION: score 3.5</pre>														
,	(xi) SE	EQUENC	E DESCR	RIPTIC	ЭМ: :	SEQ	ID N	10: 1	41:					
AGAGC(CAGGG A	ACTCGG	GTGC CT	GGGGC	CAGA	CGA	GGCC	GGC	TTCI	'CCGC	GG A	CAGO	TAGGG	60
AGAGT(GTCCŤ (GGTGT	CAGC CA	AGAAC		Ser					Ser		A AAG : Lys	113
Lys L	TG TTC eu Phe 15	ATT T	TC TTA	GGA F Gly I -10	AAA ' Lys :	TCA Ser	CTG Leu	TTT Phe	AGT Ser -5	CTT Leu	CTG Leu	GAG Glu	GCT Ala	161
ATG A' Met I I	TT TTT le Phe	GCC T Ala L	TA CTC eu Leu 5	CCA A	AAG (CCA Pro	CGG Arg 10	AAG Lys	AAC Asn	GTT Val	GCT Ala	GGT Gly 15	GAA Glu	209
ATA G	TC CTC al Leu	ATC A Ile T 20	CA GGT	GCT (GGA /	AGT Ser 25	GGA Gly	CTC Leu	GGA Gly	AGG Arg	CTC Leu 30	TTA Leu	GCC Ala	257
ITG C. Leu G	AG TTT ln Phe 35	GCC C	GG CTG	GGA 1	CT Ser	GTT Val	CTT Leu	GTT Val	CTC Leu	TGG Trp 45	GAT Asp	ATC Ile	AAT Asn	305
Lys G	AG GGG lu Gly 50													317
(2) <u>I</u>	NFORMA'	rion f	OR SEQ	ID NO	D: 1	42:								
	(1 (1	A) LEN B) TYE C) STR	CHARAC IGTH: 39 PE: NUCI RANDEDNE POLOGY:	92 bas LEIC A	se p ACID DOUB	airs	3	•						
	/: i) M	OLECUI	E TYPE	CDM	Δ									

109

(A) ORGANISM: Homo Sapiens(F) TISSUE TYPE: Large intestine

(ix) FEATURE:

(A) NAME/KEY: other (B) LOCATION: 40..94

(C) IDENTIFICATION METHOD: blastn

(D) OTHER INFORMATION: identity 92

region 1..55 id H30111

(ix) FEATURE:

(A) NAME/KEY: sig_peptide

(B) LOCATION: 15..92

(C) IDENTIFICATION METHOD: Von Heijne matrix

(D) OTHER INFORMATION: score 7.6

seq FLLLVAAPRWVVS/EM

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 142:

GGTTTCTGAG AGTC			AAA GTG AAG CAC Lys Val Lys His -20	
			TGG GTC GTG TCC Trp Val Val Ser	
CAG ATT GAG GAG Gln Ile Glu Glu 5	TCG GGC CCA Ser Gly Pro	GGC CTG GTG Gly Leu Val 10	AAG CCC TCG GAG Lys Pro Ser Glu 15	ACC CTG 146 Thr Leu
			ATC AAC ACT AAT Ile Asn Thr Asn 30	
			AAG GGT CTG GAG Lys Gly Leu Glu 45	
			TAC AAC CCG TCC Tyr Asn Pro Ser	
			AGG AAT CAG TTC Arg Asn Gln Phe 80	
CAG ATG AGT TCT Gln Met Ser Ser 85	GTG ATG GCC Val Met Ala	ACA GAC ACG Thr Asp Thr 90	GCT GTC TAT TAT Ala Val Tyr Tyr 95	TGT GCG 386 Cys Ala
AGA CAA Arg Gln 100				392

 (i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 253 base pairs (B) TYPE: NUCLEIC ACID (C) STRANDEDNESS: DOUBLE (D) TOPOLOGY: LINEAR 		
(ii) MOLECULE TYPE: CDNA		
<pre>(vi) ORIGINAL SOURCE: (A) ORGANISM: Homo Sapiens (F) TISSUE TYPE: Lung (cells)</pre>		
(ix) FEATURE: (A) NAME/KEY: other (B) LOCATION: 91253 (C) IDENTIFICATION METHOD: fasta (D) OTHER INFORMATION: identity 98.8 region 109271 id HSU73682 vrt		
(ix) FEATURE: (A) NAME/KEY: other (B) LOCATION: 220253 (C) IDENTIFICATION METHOD: blastn (D) OTHER INFORMATION: identity 94 region 134 id M78620 est		
<pre>(ix) FEATURE: (A) NAME/KEY: sig_peptide (B) LOCATION: 65220 (C) IDENTIFICATION METHOD: Von Heijne matrix (D) OTHER INFORMATION: score 6.6</pre>		
ATATGANGTG CCTGGTAGAA CTTTATTTAC ATGTTAGGAA ATTTTAACAT GAGCT	TTTCA	60
AGAA ATG GAA TTG AAA AGT CCA GAG GAA GAG GTT GTG GCA GCA CTG Met Glu Leu Lys Ser Pro Glu Glu Glu Val Val Ala Ala Leu -50 -45 -40	CCT :	109
GAA GGT ATG AGA CCA GAT TCT AAT CTT TAT GGT TTT CCA TGG GAA Glu Gly Met Arg Pro Asp Ser Asn Leu Tyr Gly Phe Pro Trp Glu -35 -30 -25	TTG :	157
GTG ATA TGT GCA GCT GTT GTT GGA TTT TTT GCT GTT CTC TTT TTT /al Ile Cys Ala Ala Val Val Gly Phe Phe Ala Val Leu Phe Phe -20 -15 -10		205
rgg AgA AgT TTT RGA TCG GTT AGG AGT CGG CTT TAT GTG GGA CGA Frp Arg Ser Phe Xaa Ser Val Arg Ser Arg Leu Tyr Val Gly Arg -5 1 5	GGG :	253

(2) INFORMATION FOR SEQ ID NO: 144:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 407 base pairs
 - (B) TYPE: NUCLEIC ACID
 - (C) STRANDEDNESS: DOUBLE
 - (D) TOPOLOGY: LINEAR
- (ii) MOLECULE TYPE: CDNA
- (vi) ORIGINAL SOURCE:
 - (A) ORGANISM: Homo Sapiens
 - (F) TISSUE TYPE: Thyroid
- (ix) FEATURE:
 - (A) NAME/KEY: other
 - (B) LOCATION: 213..303
 - (C) IDENTIFICATION METHOD: blastn
 - (D) OTHER INFORMATION: identity 100

region 141..231 id AA040646

est

- (ix) FEATURE:
 - (A) NAME/KEY: other
 - (B) LOCATION: 306..389
 - (C) IDENTIFICATION METHOD: blastn
 - (D) OTHER INFORMATION: identity 95

region 233..316

id AA040646

est

- (ix) FEATURE:
 - (A) NAME/KEY: other
 - (B) LOCATION: 144..214
 - (C) IDENTIFICATION METHOD: blastn
 - (D) OTHER INFORMATION: identity 92

region 71..141 id AA040646

IU MAUSUOSO

est

- (ix) FEATURE:
 - (A) NAME/KEY: other
 - (B) LOCATION: 78..231
 - (C) IDENTIFICATION METHOD: blastn
 - (D) OTHER INFORMATION: identity 98

region 29..182

id HUML1879

est

- (ix) FEATURE:
 - (A) NAME/KEY: other
 - (B) LOCATION: 147..287
 - (C) IDENTIFICATION METHOD: blastn

(D) OTHER INFORMATION: identity 98

region 168..308 id HSC0FB071

est

- (ix) FEATURE:
 - (A) NAME/KEY: other

112

(B)	LOCATION: 60113	
(C)	IDENTIFICATION METHOD:	blastn

(D) OTHER INFORMATION: identity 98 region 1..54 id R72047

est

1	ń	v	١	FEA	١T	ΠR	F	٠

- (A) NAME/KEY: sig_peptide
- (B) LOCATION: 177..293
- (C) IDENTIFICATION METHOD: Von Heijne matrix
- (D) OTHER INFORMATION: score 6.6

seq LALVLAWLSTYVA/DS

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 144:

CTGT	CCCI	GG '	PATA	ACAGO	T TO	CGGG	SAGAP	GCC	CGGAZ	AGAG	ACC	GACC	CT (GAACA	AGAATO	60
GCAC	SATTO	SCC 2	AGCCC	CTTTI	c co	GAC	CCTA	CGC	SAAAC	SACG	AGTO	CAG	GG (CCGT	CTGGC	120
GAGO	STCAF	LAA (CATTI	TAGTO	T GO	STCTI	TTCA	GCC	GTGG!	ACCC	TGCC	CAGC	IGC (CAGG	CC ATO	
			GAT Asp -35													227
			GGT Gly													275
			TAC Tyr													323
			GCA Ala													371
			GCA Ala 30													407

(2) INFORMATION FOR SEQ ID NO: 145:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 353 base pairs
 - (B) TYPE: NUCLEIC ACID
 - (C) STRANDEDNESS: DOUBLE
 - (D) TOPOLOGY: LINEAR
- (ii) MOLECULE TYPE: CDNA
- (vi) ORIGINAL SOURCE:
 - (A) ORGANISM: Homo Sapiens
 - (F) TISSUE TYPE: Pancreas
- (ix) FEATURE:

113

(A) NAME/KEY: other (B) LOCATION: 167..303

(C) IDENTIFICATION METHOD: blastn

(D) OTHER INFORMATION: identity 100

region 110..246

id H53379

est

(ix) FEATURE:

(A) NAME/KEY: other
(B) LOCATION: 57..171

(C) IDENTIFICATION METHOD: blastn

(D) OTHER INFORMATION: identity 97

region 1..115 id H53379

est

(ix) FEATURE:

(A) NAME/KEY: other
(B) LOCATION: 308..350

(C) IDENTIFICATION METHOD: blastn

(D) OTHER INFORMATION: identity 95

region 250..292 id H53379

...

(ix) FEATURE:

(A) NAME/KEY: other

(B) LOCATION: 64..303

(C) IDENTIFICATION METHOD: blastn

(D) OTHER INFORMATION: identity 99

region 1..240 id N87636

est

(ix) FEATURE:

(A) NAME/KEY: other

(B) LOCATION: 167..322

(C) IDENTIFICATION METHOD: blastn

(D) OTHER INFORMATION: identity 97

region 52..207

id R25125

est

(ix) FEATURE:

(A) NAME/KEY: other

(B) LOCATION: 115..171

(C) IDENTIFICATION METHOD: blastn

(D) OTHER INFORMATION: identity 96

region 1..57 id R25125

est

(ix) FEATURE:

(A) NAME/KEY: other

(B) LOCATION: 94..205

(C) IDENTIFICATION METHOD: blastn

(D) OTHER INFORMATION: identity 99

region 1..112 id W00411

est

(ix) FEATURE:

PCT/IB98/01233

		(E	3) LC	CATI ENTI	ON:	othe 201. ATION ORMAT	.302 MET	HOD:	lenti gion l WOO	ty 9		18				
	(ix	(E	A) NA B) LC C) II	ME/K CATI CENTI	ON:	othe 57 ATION DRMAT	205 1 MET	i ic	lenti gion 1 N83	ty 9						
	·	E) (C) (C)	A) NA B) LO C) II O) OT	AME/R CATI CENTI	ON: FICA INFO	sig_ 153. ATION DRMAT	.275 N MET	CHOD:	ore eq RI	5.4 LLYIC	SFLGY					
AGTT	TGGG	GCG C	TGC	ACCGO	G TO	GCGCG	CGTGC	ccc	STGAC	STCC	GGC	GCGC#	AGA (GGAGG	SAGGAG	60
AAAG	CTGF	ACC C	GCTT <i>I</i>	AGGCC	C GC	GTA	STGGT	r cgi	CGT	GTT	TTC	CTTGT	rag :	rtcgi	rggtct	120
GAGA	.CCAG	GC (CTCA	AGTGO	SA AF	ACGG	CGTC	A CC						AAC Asn		173
GAA Glu	CCC Pro	TTA Leu	CGG Arg	TTT Phe -30	CTG Leu	CCG Pro	GAT Asp	GAG Glu	GCC Ala -25	CGG Arg	AGC Ser	CTG Leu	CCC Pro	CCG Pro -20	CCC Pro	221
AAG Lys	CTG Leu	ACC Thr	GAC Asp -15	CCG Pro	CGG Arg	CTC Leu	CTC Leu	TAC Tyr -10	ATC Ile	GGC Gly	TTC Phe	TTG Leu	GGC Gly -5	TAC Tyr	TGC Cys	269
TCC Ser	GGC Gly	CTG Leu 1	ATT Ile	GAT Asp	AAC Asn	CTG Leu 5	ATC Ile	CGG Arg	CGG Arg	AGG Arg	CCG Pro 10	ATC Ile	GCG Ala	ACG Thr	GCT Ala	317
						CTA Leu										353
(2)) SE(()	QUEN A; L! B) T	CE CI ENGT: YPE:	HARACH: 1	ID DETERMINED TO THE PROPERTY OF THE PROPERTY	ISTIC ase p ACI	CS: pair: D	s							

(ii)	MOLECULE	TYPE:	CDNA

(D) TOPOLOGY: LINEAR

- (vi) ORIGINAL SOURCE:
 - (A) ORGANISM: Homo Sapiens(F) TISSUE TYPE: Thyroid
- (ix) FEATURE:
 - (A) NAME/KEY: other
 - (B) LOCATION: 53..191
 - (C) IDENTIFICATION METHOD: blastn
 - (D) OTHER INFORMATION: identity 100 region 27..165

id T96213 est

- (ix) FEATURE:
 - (A) NAME/KEY: other
 - (B) LOCATION: 40..191
 - (C) IDENTIFICATION METHOD: blastn
 - (D) OTHER INFORMATION: identity 99

region 1..152 id AA156832

est

- (ix) FEATURE:
 - (A) NAME/KEY: sig_peptide
 - (B) LOCATION: 52..123
 - (C) IDENTIFICATION METHOD: Von Heijne matrix
 - (D) OTHER INFORMATION: score 5.2

seq VGGLILWLSVGSS/GD

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 146:

TTTGGCTTTA GGGAATTACT CCATACCAGC TCTGAGATTT CCAGCTCAGC G ATG CCC 57

Met Pro

CCA GGT CCC TGG GAG AGC TGC TTC TGG GTG GGG GGC CTC ATT TTG TGG
Pro Gly Pro Trp Glu Ser Cys Phe Trp Val Gly Gly Leu Ile Leu Trp
-20 -15 -10

CTC AGC GTT GGA AGT TCA GGG GAT GCA CCT CCT ACC CCA CAG CCA AAG

Leu Ser Val Gly Ser Ser Gly Asp Ala Pro Pro Thr Pro Gln Pro Lys

-5

10

TGC GCT GAC TTC CAG AGC GCC AAC CTT TTT GAA GGC ACT CGG
Cys Ala Asp Phe Gln Ser Ala Asn Leu Phe Glu Gly Thr Arg
15
20

- (2) INFORMATION FOR SEQ ID NO: 147:
 - (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 317 base pairs
 - (B) TYPE: NUCLEIC ACID
 - (C) STRANDEDNESS: DOUBLE
 - (D) TOPOLOGY: LINEAR

(ii)	MOLECULE TYPE: CDNA	
(vi)	ORIGINAL SOURCE: (A) ORGANISM: Homo Sapiens (F) TISSUE TYPE: Lung	
(ix)	FEATURE: (A) NAME/KEY: other (B) LOCATION: 248317 (C) IDENTIFICATION METHOD: blastn (D) OTHER INFORMATION: identity 95 region 170 id T27030 est	
(ix)	FEATURE: (A) NAME/KEY: other (B) LOCATION: 252317 (C) IDENTIFICATION METHOD: blastn (D) OTHER INFORMATION: identity 95	
(ix)	FEATURE: (A) NAME/KEY: sig_peptide (B) LOCATION: 210251 (C) IDENTIFICATION METHOD: Von Heijne matrix (D) OTHER INFORMATION: score 5 seq CARALLLACSSRG/RH	
(xi)	SEQUENCE DESCRIPTION: SEQ ID NO: 147:	
ACACTTCCA	G CTGCTCCCTC CAAAGTTTGT CCTAGTTTCC CCCTTCCCAG CTCTCCCCAT	60
rcgcagsct(C TTCCATCTTC AACTCTTCTT CCCKGCTVAC STCKTMCCTC VTCCGATGGT 1	20
CTCTCCTTG	G CTTCKCTCCA CSCCGCTTGC YTCTTCTCTA GTCTTTCCCT GGCCCTGGCA 1	.80
TAGTCTCC	T TACCCTGTGC CCTGTCCCA ATG TGT GCC CGG GCT TTG CTC CTT 2 Met Cys Ala Arg Ala Leu Leu Leu -10	:33
GCG TGC A Ala Cys S -5	GT TCG AGG GGM AGA CAT CGT TTG GCN TGT CAG TGT TCA ACC 2 er Ser Arg Gly Arg His Arg Leu Ala Cys Gln Cys Ser Thr 1 5 10	81
ICA GCA A Ser Ala T	CG CCA TCA TGG GCA GCG GCA TCC TGG GGC hr Pro Ser Trp Ala Ala Ala Ser Trp Gly 15 20	117
	MATION FOR SEQ ID NO: 148: SEQUENCE CHARACTERISTICS: (A) LENGTH: 367 base pairs (B) TYPE: NUCLEIC ACID (C) STRANDEDNESS: DOUBLE (D) TOPOLOGY: LINEAR	

(ii) MOLECULE TYPE: CDNA (vi) ORIGINAL SOURCE: (A) ORGANISM: Homo Sapiens (F) TISSUE TYPE: Lung (cells) (ix) FEATURE: (A) NAME/KEY: other (B) LOCATION: 200..369 (C) IDENTIFICATION METHOD: blastn (D) OTHER INFORMATION: identity 98 region 109..278 id H38087 est (ix) FEATURE: (A) NAME/KEY: other (B) LOCATION: 58..129 (C) IDENTIFICATION METHOD: blastn (D) OTHER INFORMATION: identity 94 region 41..112 id H38087 est (ix) FEATURE: (A) NAME/KEY: other (B) LOCATION: 200..360 (C) IDENTIFICATION METHOD: blastn (D) OTHER INFORMATION: identity 98 region 63..223 id R85713 est (ix) FEATURE: (A) NAME/KEY: other (B) LOCATION: 64..129 (C) IDENTIFICATION METHOD: blastn (D) OTHER INFORMATION: identity 100 region 1..66 id R85713 est (ix) FEATURE: (A) NAME/KEY: sig_peptide (B) LOCATION: 44..181 (C) IDENTIFICATION METHOD: Von Heijne matrix (D) OTHER INFORMATION: score 4.7 seq IICCVFLLLAIVG/YV (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 148: AAGAGAGAGC GCGGGGGCC GCCGGGGCCTG GTCGCCTGCA GGG ATG GGG GAC GAG Met Gly Asp Glu -45 CGG CCC CAC TAC TAC GGG AAA CAC GGA ACG CCA CAG AAG TAT GAT CCC Arg Pro His Tyr Tyr Gly Lys His Gly Thr Pro Gln Lys Tyr Asp Pro

-35

118

•								
TTC Phe -25								151
GTG Val								199
GCC Ala								247
CGG Arg								295
TAT Tyr 40								343
CTG Leu	-	 	 					367

(2) INFORMATION FOR SEQ ID NO: 149:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 327 base pairs
 - (B) TYPE: NUCLEIC ACID
 - (C) STRANDEDNESS: DOUBLE
 - (D) TOPOLOGY: LINEAR
- (ii) MOLECULE TYPE: CDNA
- (vi) ORIGINAL SOURCE:
 - (A) ORGANISM: Homo Sapiens
 - (F) TISSUE TYPE: Lung (cells)
- (ix) FEATURE:
 - (A) NAME/KEY: other
 - (B) LOCATION: 85..328
 - (C) IDENTIFICATION METHOD: fasta
 - (D) OTHER INFORMATION: identity 98.8 region 1..244 id HSU78678

vrt

- (ix) FEATURE:
 - (A) NAME/KEY: other
 - (B) LOCATION: 85..328
 - (C) IDENTIFICATION METHOD: blastn
 - (D) OTHER INFORMATION: identity 97

region 20..263

id N41898

est

- (ix) FEATURE:
 - (A) NAME/KEY: other
 - (B) LOCATION: 98..328

(C) IDENTIFICATION METHOD: blastn (D) OTHER INFORMATION: identity 95 region 38268 id H69272 est	
(ix) FEATURE: (A) NAME/KEY: sig_peptide (B) LOCATION: 85132 (C) IDENTIFICATION METHOD: Von Heijne matrix (D) OTHER INFORMATION: score 4.5 seq RLLLRRFLASVIS/RK (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 149:	
CCCCCGAGG GAAGTGACGA CAGGCGTGCC CTTGACAGGC AGGGAGGGCT AGGCTGTGCA	60
TCCCTCCGCT CGCATTGCAG GGAG ATG GCT CAG CGA CTT CTT CTG AGG AGG Met Ala Gln Arg Leu Leu Arg Arg -15 -10	111
TTC CTG GCC TCT GTC ATC TCC AGG AAG CCC TCT CAG GGT CAG TGG CCA Phe Leu Ala Ser Val Ile Ser Arg Lys Pro Ser Gln Gly Gln Trp Pro -5 1 5	159
CCC CTC ACT TCC AGA GCC CTG CAG ACC CCA CAA TGC AGT CCT GGT GGC Pro Leu Thr Ser Arg Ala Leu Gln Thr Pro Gln Cys Ser Pro Gly Gly 10 15 20 25	207
CTG ACT GTA ACA CCC AAC CCA GCC CGG ACA ATA TAC ACC ACG AGG ATC Leu Thr Val Thr Pro Asn Pro Ala Arg Thr Ile Tyr Thr Thr Arg Ile 30 35 40	255
TCC TTG ACA ACC TTT AAT ATC CAG GAT GGA CCT GAC TTT CAA GAC CGA Ser Leu Thr Thr Phe Asn Ile Gln Asp Gly Pro Asp Phe Gln Asp Arg 45 50 55	303
GTG GTC AAC AGT GAG ACA CCG GCG Val Val Asn Ser Glu Thr Pro Ala 60 65	327
(2) INFORMATION FOR SEQ ID NO: 150: (i) SEQUENCE CHARACTERISTICS:	
(A) LENGTH: 378 base pairs(B) TYPE: NUCLEIC ACID(C) STRANDEDNESS: DOUBLE(D) TOPOLOGY: LINEAR	
(ii) MOLECULE TYPE: CDNA	
<pre>(vi) ORIGINAL SOURCE: (A) ORGANISM: Homo Sapiens (F) TISSUE TYPE: Lung (cells)</pre>	·
(im) FEATURE: (A) NAME/KEY: other (B) LOCATION: 203340	

(C) IDENTIFICATION METHOD: blastn

120

(D) OTHER INFORMATION: identity 91 region 33..170 id T50032 (ix) FEATURE: (A) NAME/KEY: other (B) LOCATION: 329..376 (C) IDENTIFICATION METHOD: blastn (D) OTHER INFORMATION: identity 100 region 160..207 id T50032 (ix) FEATURE: (A) NAME/KEY: other (B) LOCATION: 170..210 (C) IDENTIFICATION METHOD: blastn (D) OTHER INFORMATION: identity 97 region 1..41 id T50032 est (ix) FEATURE: (A) NAME/KEY: sig_peptide (B) LOCATION: 109..177 (C) IDENTIFICATION METHOD: Von Heijne matrix (D) OTHER INFORMATION: score 4.4 seq QFILLGTTSVVTA/AL (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 150: AAGGTTAGGC CGGGGGGGTG CGGTCCTGGT CGGAAGGAGG HSBAKWSYCG MGVVNGTCAC CAGGCCTATC CTTGGCGCCA CAGTCGGCCA CCGGGGCTCG CCGCCGTC ATG GAG AGC 117 Met Glu Ser GGA GGG CGG CCC TCG CTG TGC CAG TTC ATC CTC CTG GGC ACC ACC TCT 165 Gly Gly Arg Pro Ser Leu Cys Gln Phe Ile Leu Leu Gly Thr Thr Ser -20 -15GTG GTC ACC GCC GCC CTG TAC TCC GTG TAC CGG CAG AAG GCC CGG GTC 213 Val Val Thr Ala Ala Leu Tyr Ser Val Tyr Arg Gln Lys Ala Arg Val TCC CAA GAG CTC AAG GGA GCT AAA AAA GTT CAT TTG GGT GAA GAT TTA 261 Ser Gln Glu Leu Lys Gly Ala Lys Lys Val His Leu Gly Glu Asp Leu AAG AGT ATT CTT TCA GAA GCT CCA GGA AAA TGC GTS CCT DAT GCT GTY 309 Lys Ser Ile Leu Ser Glu Ala Pro Gly Lys Cys Val Pro Xaa Ala Val ATA GAA GGA GCT GTG CGG TCT GTT AAA GAA ACG CTT AAC AGC CAG TTT 357 Ile Glu Gly Ala Val Arg Ser Val Lys Glu Thr Leu Asn Ser Gln Phe 50 GTG GAA AAC TGC AAG GGG GAS 378 Val Glu Ash Cys Lys Gly Xaa

PCT/IB98/01233 WO 99/06439

121

. 65

(2) INFORMATION FOR SEQ ID NO: 151:	
(i) SEQUENCE CHARACTERISTICS:(A) LENGTH: 267 base pairs(B) TYPE: NUCLEIC ACID(C) STRANDEDNESS: DOUBLE(D) TOPOLOGY: LINEAR	
(ii) MOLECULE TYPE: CDNA	
<pre>(vi) ORIGINAL SOURCE: (A) ORGANISM: Homo Sapiens (F) TISSUE TYPE: Liver</pre>	
(ix) FEATURE: (A) NAME/KEY: other (B) LOCATION: 186261 (C) IDENTIFICATION METHOD: blastn (D) OTHER INFORMATION: identity 98 region 2095 id T48277 est	
<pre>(ix) FEATURE: (A) NAME/KEY: other (B) LOCATION: complement(185261) (C) IDENTIFICATION METHOD: blastn (D) OTHER INFORMATION: identity 100</pre>	
(ix) FEATURE: (A) NAME/KEY: sig_peptide (B) LOCATION: 49120 (C) IDENTIFICATION METHOD: Von Heijne matrix (D) OTHER INFORMATION: score 16.4 seq LLLLLLLASLTSG/SV (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 151:	
AAGAGCAAGC TCAAGACCCA GCAGTGGGAC AGCCAGACAG ACGGCACG ATG GCA CTG	57
AGC TCC CAG ATC TGG GCC GCT TGC CTC CTC CTC CTC CTC GCC Ser Ser Gln Ile Trp Ala Ala Cys Leu Leu Leu Leu Leu Leu Leu Leu Ala -20 -15 -10	105
AGC CTG ACC AGT GGC TCT GTT TTC CCA CAA CAG ACG GGA CAA CTT GCA Ser Leu Thr Ser Gly Ser Val Phe Pro Gln Gln Thr Gly Gln Leu Ala -5 1 5 10	153
GAR CTG CAA CCC CAG GAC AGA GCT GGA GCC AGG GCC AGC TGG ATG CCC Glu Leu Gln Pro Gln Asp Arg Ala Gly Ala Arg Ala Ser Trp Met Pro	201

122

ATG TTC CAG AGG CGA AGG CGA GAC CAC CAC TTC CCC ATC TGC ATT

Met Phe Gln Arg Arg Arg Arg Arg Asp Thr His Phe Pro Ile'Cys Ile

30 35 40

TTC TGC TGC GGC CCT GGG Phe Cys Cys Gly Pro Gly 45 267

(2) INFORMATION FOR SEQ ID NO: 152:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 343 base pairs
 - (B) TYPE: NUCLEIC ACID
 - (C) STRANDEDNESS: DOUBLE
 - (D) TOPOLOGY: LINEAR
- (ii) MOLECULE TYPE: CDNA
- (vi) ORIGINAL SOURCE:
 - (A) ORGANISM: Homo Sapiens
 - (F) TISSUE TYPE: Large intestine
- (ix) FEATURE:
 - (A) NAME/KEY: other
 - (B) LOCATION: 69..297
 - (C) IDENTIFICATION METHOD: blastn
 - (D) OTHER INFORMATION: identity 96

region 25..253 id C16912

est

- (ix) FEATURE:
 - (A) NAME/KEY: other
 - (B) LOCATION: 181..269
 - (C) IDENTIFICATION METHOD: blastn
 - (D) OTHER INFORMATION: identity 94

region 64..152

id T68684

est

- (ix) FEATURE:
 - (A) NAME/KEY: other
 - (B) LOCATION: 133..173
 - (C) IDENTIFICATION METHOD: blastn
 - (D) OTHER INFORMATION: identity 90

region 18..58

id T68684

est

- (ix) FEATURE:
 - (A) NAME/KEY: sig_peptide
 - (B) LOCATION: 107..175
 - (C) IDENTIFICATION METHOD: Von Heijne matrix
 - (D) OTHER INFORMATION: score 14.4

seq LGLLLFLLPGSLG/AE

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 152:

ACAC	GATO	STG A	AGAGA	AGGA	AC TO	GGGT	CTCC	AG1	CAC	GGA	GCC	GGAC	SCC (GCCI	AGGGC	60
GCA	GAGN	NGA P	AGGGF	AGCG <i>I</i>	AG GC	CTGA	AGGG <i>I</i>	A ACC	STCG1	CCT	CTC			GGG (115
														CTT Leu		163
														CTT Leu		211
														TCC Ser		259
														G1y		307
			GTG Val													343

(2) INFORMATION FOR SEQ ID NO: 153:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 284 base pairs
 - (B) TYPE: NUCLEIC ACID
 - (C) STRANDEDNESS: DOUBLE
 - (D) TOPOLOGY: LINEAR
- (ii) MOLECULE TYPE: CDNA
- (vi) ORIGINAL SOURCE:
 - (A) ORGANISM: Homo Sapiens
 - (F) TISSUE TYPE: Liver
- (ix) FEATURE:
 - (A) NAME/KEY: other
 - (B) LOCATION: 25..177
 - (C) IDENTIFICATION METHOD: blastn
 - (D) OTHER INFORMATION: identity 100 region 1..153 id T60354

est

- (ix) FEATURE:
 - (A) NAME/KEY: sig_peptide
 - (B) LOCATION: 132..185
 - (C) IDENTIFICATION METHOD: Von Heijne matrix
 - (D) OTHER INFORMATION: score 10.3
 - seq SLLLSVLLAQVWL/VP
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 153:

. 124	
AGGACCCAAG GCCACACACT GGAAGTCTTG CAGCTGAAGG GAGGCACTCC TTGGCCTCCG	120
CAGCCGATCA C ATG AAG GTG GTG CCA AGT CTC CTG CTC TCC GTC CTC CTG Met Lys Val Val Pro Ser Leu Leu Ser Val Leu Leu -15 -10	170
GCA CAG GTG TGG CTG GTA CCC GGC TTG GCC CCC AGT CCT CAG TCG CCA Ala Gln Val Trp Leu Val Pro Gly Leu Ala Pro Ser Pro Gln Ser Pro -5 1 5 10	218
GAG ACC CCA GCC CCT CAG AAC CAG ACC AGC AGG GTA GTG CAG GCT CCC Glu Thr Pro Ala Pro Gln Asn Gln Thr Ser Arg Val Val Gln Ala Pro 15 20 25	266
AGG GAG GAA GAG GAA TGG Arg Glu Glu Glu Trp 30	284
(2) INFORMATION FOR SEQ ID NO: 154:	
(i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 607 base pairs (B) TYPE: NUCLEIC ACID (C) STRANDEDNESS: DOUBLE (D) TOPOLOGY: LINEAR (ii) MOLECULE TYPE: CDNA (vi) ORIGINAL SOURCE: (A) ORGANISM: Homo Sapiens (F) TISSUE TYPE: Pancreas (ix) FEATURE: (A) NAME/KEY: other (B) LOCATION: 23517 (C) IDENTIFICATION METHOD: fasta (D) OTHER INFORMATION: identity 96.4 region 1495 id S82198 vrt	
(ix) FEATURE: (A) NAME/KEY: other (B) LOCATION: 155266 (C) IDENTIFICATION METHOD: blastn (D) OTHER INFORMATION: identity 97 region 118229 id T47757 est	
<pre>(ix) FEATURE: (A) NAME/KEY: other (B) LOCATION: 64155 (C) IDENTIFICATION METHOD: blastn (D) OTHER INFORMATION: identity 95</pre>	

id T47757 est

(ix) FEATURE:

(A) NAME/KEY: sig_peptide

(B) LOCATION: 23..70

- (C) IDENTIFICATION METHOD: Von Heijne matrix
- (D) OTHER INFORMATION: score 9.5

seq ITVLAALLACASS/CG

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 154:

GGCCAGTCCT GAGCACCTAA CC ATG TTG AGC ATC ACT GTC CTC GCT GCG CTC Met Leu Ser Ile Thr Val Leu Ala Ala Leu -15												
TTG GCC T Leu Ala C	TGT GCC Cys Ala	TCC AGC Ser Ser	TGT GGG Cys Gly 1	GTG CC Val Pr	CC AGC co Ser 5	TTC CCG Phe Pro	CCC AAC Pro Asn	CTA Leu 10	100			
TCC GCC C Ser Ala A				Asp Al				Pro	149			
TGG CAG A	ATC TCC Ile Ser 30	CTC CAG Leu Gln	TAC CTC Tyr Leu	AAG AA Lys As 35	AC GAC sn Asp	ACG TGG Thr Trp	AGG CAT Arg His 40	ACG Thr	196			
TGT GGC (Cys Gly (244			
TGC ATC A Cys Ile S 60									292			
CTG GAG C Leu Glu V 75									340			
ATC CAC (Ala Xa				Ile	388			
GCC CTC A									436			
GTG GCS 1									484			
TGC TAT (Cys Tyr 1									532			
SCA CTG A Naa Leu A 155	AAC TGC Asn Cys	CAG TTG Gln Leu 160	GAG AAC Glu Asn	GGT TO	CC TGG er Trp 165	GAG GTG Glu Val	TTT GGC	WTC Xaa 170	530			
GTC AGC Val Ser									607			

PCT/IB98/01233 WO 99/06439 126

(2)	INFO	RMAT	ION	F OR	SEQ	ID N	10: 1	.55:								
	(i)	(B (C) LE) TY) ST	NGTH PE: RAND	IARAC I: 39 NUCI DEDNE DGY:	2 ba EIC SS:	se p ACID DOUE	airs)	3							
	(ii) MO	LECU	LE I	YPE:	CDN	IA									
	(vi) OR	GANI	OURC SM: TYP	Homo			i							
	(ix	(B (C) NA) LC) IE	ME/R CATI ENTI	CEY: CON: FICA INFO	156.	.267 MET	HOD:	lenti gion i T47	ty 9	97 122	9				
	(ix	(B (C) NA) LC) IC	ME/F CATI ENTI	(EY: ON: FICA INFO	65 TION	156 MET	i io	lenti gion i T47	ty 9 26.	94 . 117	,				
	·	(B (C) NA) LC) IE) OI	ME/F CATI CENTI CHER	(EY: ION: IFICA INFO	24 ATION ORMAT	71 NMET	HOD: se	ore eq IT	9.5 YLA	LLAC	,				
AAGO	CAGI	CC T	GAGC	CACCI	OA AT			eu Gl					eu Al		CG CTC la Leu	53
		TGT Cys														101
		CGA Arg														149
TGG Trp	CAG Gln	ATC Ile	TCC Ser 30	CTC Leu	CAG Gln	TAC Tyr	CTC Leu	AAG Lys 35	AAC Asn	GAC Asp	ACG Thr	TGG Trp	AGG Arg 40	CAT His	ACG Thr	197

		TTG Leu						245
		ACC Thr						293
		GAC Asp						341
		AGG Arg 95						389
GCC Ala								392

(2) INFORMATION FOR SEQ ID NO: 156:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 240 base pairs
 - (B) TYPE: NUCLEIC ACID
 - (C) STRANDEDNESS: DOUBLE
 - (D) TOPOLOGY: LINEAR
- (ii) MOLECULE TYPE: CDNA
- (vi) ORIGINAL SOURCE:
 - (A) ORGANISM: Homo Sapiens
 - (F) TISSUE TYPE: Colon
- (ix) FEATURE:
 - (A) NAME/KEY: other
 - (B) LOCATION: 102..240
 - (C) IDENTIFICATION METHOD: blastn
 - (D) OTHER INFORMATION: identity 97

region 83..221 id H83276

est

- (ix) FEATURE:
 - (A) NAME/KEY: other
 - (B) LOCATION: 66..240
 - (C) IDENTIFICATION METHOD: blastn
 - (D) OTHER INFORMATION: identity 97

region 23..197 id H51676

est

- (ix) FEATURE:
 - (A) NAME/KEY: other
 - (B) LOCATION: 91..240
 - (C) IDENTIFICATION METHOD: blastn
 - (D) OTHER INFORMATION: identity 95 region 71..220 id AA007645

est

128

(ix) FEATURE:

(A) NAME/KEY: other

(B) LOCATION: 16..68

(C) IDENTIFICATION METHOD: blastn

(D) OTHER INFORMATION: identity 98

region 1..53 id AA007645

est

(ix) FEATURE:

(A) NAME/KEY: other (B) LOCATION: 64..240

(C) IDENTIFICATION METHOD: blastn

(D) OTHER INFORMATION: identity 99

region 74..250 id W95024

.u w 93024

est

(ix) FEATURE:

(A) NAME/KEY: other

(B) LOCATION: 119..161

(C) IDENTIFICATION METHOD: blastn

(D) OTHER INFORMATION: identity 97

region 417..459 id AA149660

est

(ix) FEATURE:

(A) NAME/KEY: other

(B) LOCATION: 19..52

(C) IDENTIFICATION METHOD: blastn

(D) OTHER INFORMATION: identity 100

region 314..347 id AA149660

est

(ix) FEATURE:

(A) NAME/KEY: other

(B) LOCATION: 66..95

(C) IDENTIFICATION METHOD: blastn

(D) OTHER INFORMATION: identity 93

region 361..390 id AA149660

est

(ix) FEATURE:

(A) NAME/KEY: sig_peptide

(B) LOCATION: 85..198

(C) IDENTIFICATION METHOD: Von Heijne matrix

(D) OTHER INFORMATION: score 8.8

seq LLLPLLSLPVTTP/WT

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 156:

AGCACTCTAG TTCAAGAGTG AAAAGTCTCA CCCAGGAGGA ACAGCCCTCC TTGAAGCAAT

111

60

GGCAGGGCCA GCAGGGAGGT GGGC ATG GCA GGG AAT GGA GAG AGT GAG CCA Met Ala Gly Asn Gly Glu Ser Glu Pro

129

-35 -30

GAC AGA CTT CAC CTC CTT ACT GGA CAC AGG GTC AAG GGC GAG TTT CAA 159
Asp Arg Leu His Leu Leu Thr Gly His Arg Val Lys Gly Glu Phe Gln
-25 -20 -15

TTG CTG CTC CCT TTA CTT TCT CTA CCT GTG ACT ACT CCC TGG ACC AAT

Leu Leu Leu Pro Leu Leu Ser Leu Pro Val Thr Thr Pro Trp Thr Asn

-10

-5

CCT GAG GAG GGC ACA TTT TCC AGA AGC CAC GGG
Pro Glu Glu Gly Thr Phe Ser Arg Ser His Gly
5 10

(2) INFORMATION FOR SEQ ID NO: 157:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 320 base pairs

- (B) TYPE: NUCLEIC ACID
- (C) STRANDEDNESS: DOUBLE
- (D) TOPOLOGY: LINEAR
- (ii) MOLECULE TYPE: CDNA
- (vi) ORIGINAL SOURCE:
 - (A) ORGANISM: Homo Sapiens
 - (F) TISSUE TYPE: Large intestine
- (ix) FEATURE:
 - (A) NAME/KEY: other
 - (B) LOCATION: complement (241..322)
 - (C) IDENTIFICATION METHOD: blastn
 - (D) OTHER INFORMATION: identity 100 region 27..108 id HSBC2H071

est

- (ix) FEATURE:
 - (A) NAME/KEY: other
 - (B) LOCATION: complement(241..322)
 - (C) IDENTIFICATION METHOD: blastn
 - (D) OTHER INFORMATION: identity 100

region 328..409

id N27679

est

- (ix) FEATURE:
 - (A) NAME/KEY: other
 - (B) LOCATION: complement(241..322)
 - (C) IDENTIFICATION METHOD: blastn
 - (D) OTHER INFORMATION: identity 98

region 144..225

id W28299

est

- (ix) FEATURE:
 - (A) NAME/KEY: other
 - (B) LOCATION: complement(254..322)
 - (C) IDENTIFICATION METHOD: blastn

		([)) O:	THER	INF	ORMA'	rion	re			95 920	57				
	(i)	(E	A) NA B) LO C) II	AME/E DCATI DENTI	ION: IFIC		olemo	re	: bla dent:	astn ity 9 n 111		54				
		(B	A) N# B) L(C) II D) OT	AME/I DCATI DENTI THER	ION: IFICA INFO	ORMA?	.92 N ME'	THOD:	ore eq L	8.7 WLVI	LLLLI	matı PTLK,				
TTT		CAA C										rctg <i>i</i>		ATG (56
TGG Trp	TGG Trp	CTA Leu -10	GTG Val	CTC Leu	CTA Leu	CTC Leu	CTA Leu -5	CCT Pro	ACA Thr	TTA Leu	AAA Lys	TCT Ser 1	GTT Val	TTT Phe	TGT Cys	104
		GTA Val														152
TGG Trp	CTC Leu	TGG Trp	CCC Pro	AAA Lys 25	CCT Pro	GAC Asp	CTT Leu	CAC His	TCT Ser 30	GGA Gly	ACG Thr	AGA Arg	ACA Thr	GAG Glu 35	GTT Val	200
		CAC His														248
CCT Pro	CTC Leu	GCT Ala 55	GGA Gly	GCA Ala	GTG Val	CCC Pro	TCA Ser 60	Pro	ACT Thr	GTC Val	TCA Ser	CGT Arg 65	CTG Leu	GAG Glu	GCA Ala	296
		CGG Arg														320
(2)		(E	QUENCA) L! B) T'	CE CI ENGTI YPE: IRANI	HARAG H: 2' NUC: DEDNI	CTER	ISTICASE ACI	CS: pair: D	5							

(ii)	MOLECULE	TYPE:	CDNA
------	----------	-------	------

- (vi) ORIGINAL SOURCE:
 - (A) ORGANISM: Homo Sapiens
 - (F) TISSUE TYPE: Liver
- (ix) FEATURE:
 - (A) NAME/KEY: other
 - (B) LOCATION: 114..200
 - (C) IDENTIFICATION METHOD: blastn

(D) OTHER INFORMATION: identity 100

region 51..137 id AA120830

est

- (ix) FEATURE:
 - (A) NAME/KEY: other
 - (B) LOCATION: 197..252
 - (C) IDENTIFICATION METHOD: blastn

(D) OTHER INFORMATION: identity 91

region 135..190 id AA120830

est

- (ix) FEATURE:
 - (A) NAME/KEY: sig_peptide
 - (B) LOCATION: 98..184
 - (C) IDENTIFICATION METHOD: Von Heijne matrix
 - (D) OTHER INFORMATION: score 8.5

seq LLGLLMAACFTFC/LS

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 158:

AGCGTTTGCG CAGGGGGAGC TGGTCGCCGC CGCGGCCGCC TGGAATTGTG GGAGTNGTGT 60

CTGCCACTCG GCTGCCGGAG GCGGAARGTC CGTGACT ATG GCT CCC CAR AGC CTG 115

Met Ala Pro Gln Ser Leu

-25

CCT TCA TCT AGG ATG GCT CCT CTG GGC ATG CTG CTT GGG CTG CTG ATG

Pro Ser Ser Arg Met Ala Pro Leu Gly Met Leu Leu Gly Leu Leu Met

-20

-15

GCC GCC TGC TTC ACC TTC TGC CTC AGT CAT CAG AAC CTG AAG GAG TAT 211

Ala Ala Cys Phe Thr Phe Cys Leu Ser His Gln Asn Leu Lys Glu Tyr
-5 1 5

GCC CTG ACC AAC CCA GAK AAG ASC AGC ACC AAA GAA ACG GAA GGG
Ala Leu Thr Asn Pro Xaa Lys Xaa Ser Thr Lys Glu Thr Glu Gly
10 15 20

- (2) INFORMATION FOR SEQ ID NO: 159:
 - (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 278 base pairs
 - (B) TYPE: NUCLEIC ACID '
 - (C) STRANDEDNESS: DOUBLE
 - (D) TOPOLOGY: LINEAR

(vi) ORIGINAL SOURCE:

WO 99/06439 PCT/IB98/01233

	(11)	MOLEC	TPE 1	YPE:	CDN	IA									
	(vi)	ORIGII (A) OI (F) T	RGANI	SM:	Homo		piens	;							
	(ix)	FEATU (A) NA (B) La (C) II (D) O	AME/F OCATI DENTI	ON: FICE	25 TION	247 I ME:	: ic	ienti egior i T94	ty 9						
		FEATU (A) N. (B) L (C) I (D) O	AME/F OCATI DENTI THER	ON: FICA INFO	69 ATION DRMAT	137 N MET	THOD: se	ore eq LI	8.1 LVTV	'SSNI					
A.A.T.(CCAGAA	T ACAT	TTCC	AW C	\AGA(CAC'	r ggo	CCAAC	STCA	GCT	CTTC	CTG A	AGAG <i>F</i>	AGTCTC	60
TAG		ATG AT Met Me			s Sei					ı Cys					110
		CT TCC er Ser													158
	Gln T	CA CTC hr Leu 10													206
		AA GAA lu Glu													254
		TT CAT													278
(2)		SEQUEN (A) I (B) I (C) S (D) I	CE C ENGT YPE:	HARA H: 2 NUC DEDN	CTER 90 b LEIC ESS:	ISTI ase ACI DOU	CS: pair D	s						÷	
	(ii)	MOLEC	ULE	TYPE	: CD	NA									

(A) ORGANISM: Homo Sapiens (F) TISSUE TYPE: Liver

(ix) FEATURE:

(A) NAME/KEY: other

(B) LOCATION: 15..203 (C) IDENTIFICATION METHOD: blastn

(D) OTHER INFORMATION: identity 99

region 6..194 id T55097

est

(ix) FEATURE:

(A) NAME/KEY: other

(B) LOCATION: 199..262

(C) IDENTIFICATION METHOD: blastn

(D) OTHER INFORMATION: identity 93

region 189..252

id T55097

est

(ix) FEATURE:

(A) NAME/KEY: other

(B) LOCATION: 22..203

(C) IDENTIFICATION METHOD: blastn

(D) OTHER INFORMATION: identity 99

region 10..191

id W86026

est

(ix) FEATURE:

(A) NAME/KEY: other

(B) LOCATION: 239..283

(C) IDENTIFICATION METHOD: blastn

(D) OTHER INFORMATION: identity 91

region 225..269

id W86026

est

(ix) FEATURE:

(A) NAME/KEY: other

(B) LOCATION: 30..283

(C) IDENTIFICATION METHOD: blastn

(D) OTHER INFORMATION: identity 96

region 1..254

id H49257

est

(ix) FEATURE:

(A) NAME/KEY: other

(B) LOCATION: 42..283

(C) IDENTIFICATION METHOD: blastn

(D) OTHER INFORMATION: identity 98

region 1..242

id W84512

est

(ix) FEATURE:

(A) NAME/KEY: other

(B) LOCATION: 35..214

(C) IDENTIFICATION METHOD: blastn

		(1	D) O	rher	INFO	ORMA'	rion	re	denti egior d T69 st	ı Ī.,						
		(1 (1 (1	B) L(C) II D) O	AME/I DCAT: DENT: THER	KEY: ION: IFIC! INFO	33. ATION CAMAN	.110 N ME' TION	THOD:	core	7.6 [TLCI	LVCI	matı VANA,				
	·	,	_													
AAC	CAC	CGC (CTGT	CCTT	CC TO	GACA	CCTC	A CC				GGA Gly				53
CGC Arg	TGT Cys	GTG Val	GGG Gly	CTC Leu -15	TCC Ser	CTC Leu	ATT Ile	ACC Thr	CTC Leu -10	TGC Cys	CTC Leu	GTC Val	TGC Cys	ATT Ile -5	GTG Val	101
GCC Ala	AAC Asn	GCC Ala	CTC Leu 1	CTG Leu	CTG Leu	GTA Val	CCT Pro 5	AAT Asn	GGG Gly	GAG Glu	ACC Thr	TCC Ser 10	TGG Trp	ACC Thr	AAC Asn	149
ACC Thr	AAC Asn 15	CAT His	CTC Leu	AGC Ser	TTG Leu	CAA Gln 20	GTC Val	TGG Trp	CTC Leu	ATG Met	GGC Gly 25	GGC Gly	TTC Phe	ATT Ile	GGC Gly	197
GGG Gly 30	GGC Gly	CTA Leu	ATG Met	GTA Val	CTG Leu 35	TGT Cys	CCG Pro	GGG Gly	ATT	GCA Ala 40	GCC Ala	GTT Val	CGG Arg	GCA Ala	GGG Gly 45	245
					GGT Gly											290
(2)	INF	ORMA	TION	FOR	SEQ	ID I	NO:	161:								
	(i	(. ()	A) L: B) T' C) S'	ENGT YPE: TRAN	HARAG H: 1! NUC! DEDN! OGY:	97 b LEIC ESS:	ase j ACI DOU	pair D	s							
	(i	i) M	OLEC	ULE	TYPE	: CD	NA									
	(v	i) O	RIGI	NAL	SOUR	CE:										

(A) ORGANISM: Homo Sapiens(F) TISSUE TYPE: Lung (cells)

(C) IDENTIFICATION METHOD: blastn
(D) OTHER INFORMATION: identity 95

region 65..179

(A) NAME/KEY: other(B) LOCATION: 78..192

(ix) FEATURE:

135

id T60981 est

(ix)	FEATURE:

(A) NAME/KEY: other

(B) LOCATION: 13..75

(C) IDENTIFICATION METHOD: blastn

(D) OTHER INFORMATION: identity 92

region 1..63 id T60981 est

(ix) FEATURE:

(A) NAME/KEY: sig_peptide

(B) LOCATION: 90..158

(C) IDENTIFICATION METHOD: Von Heijne matrix

(D) OTHER INFORMATION: score 7.5

seq LVLLLTLPLHLMA/LL

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 161:

ACAGCGGGAG GGGACGCCAG CGCCTGCAGA GGCTGAGCAG GGAAAAAGCC AGTDCCCCAG 60

CGGAAGCACA GCTCABBAGC TGGTCTGCC ATG GAC ATC CTG GTC CCA CTC CTG

Met Asp Ile Leu Val Pro Leu Leu

-20

CAG CTG CTG GTG CTG CTT CTT ACC CTG CCC CTG CAC CTC ATG GCT CTG
Gln Leu Leu Val Leu Leu Thr Leu Pro Leu His Leu Met Ala Leu
-15
-5
161

CTG GGC TGC TGG CAG CCC CTG TGC AAA AGC TTT GGG
Leu Gly Cys Trp Gln Pro Leu Cys Lys Ser Phe Gly
5

(2) INFORMATION FOR SEQ ID NO: 162:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 175 base pairs
 - (B) TYPE: NUCLEIC ACID
 - (C) STRANDEDNESS: DOUBLE
 - (D) TOPOLOGY: LINEAR
- (ii) MOLECULE TYPE: CDNA
- (vi) ORIGINAL SOURCE:
 - (A) ORGANISM: Homo Sapiens
 - (F) TISSUE TYPE: Liver

(ix) FEATURE:

- (A) NAME/KEY: other
- (B) LOCATION: complement(112..142)
- (C) IDENTIFICATION METHOD: blastn
- (D) OTHER INFORMATION: identity 93 region 187..217 id AA146544 est

(ix) FEATURE:	
(A) NAME/KEY: other	
(B) LOCATION: complement(112142)	
(C) IDENTIFICATION METHOD: blastn	
(D) OTHER INFORMATION: identity 93	-
region 187217	
id AA146541 est	
650	
(ix) FEATURE:	
(A) NAME/KEY: sig_peptide	
(B) LOCATION: 107151	
(C) IDENTIFICATION METHOD: Von Heijne matrix	
(D) OTHER INFORMATION: score 7.5	
seq FLVLFSFFNIALC/AP	
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 162:	
ATTCTTTTTT CGGACTTGGA AGTTTTTATG ATTAGTGTCT TTCTGGCAGA CTTCAGTGAC	60
MEMBERS CARCENARD COMPANY COMPANY MEMBERS COMPANY COMP	
TTTTCTAGAA GCAGGTAAGA CCTAGCATGC TTTGTCCTGT GCGTTG ATG CCT TTT	115
Met Pro Phe -15	
-13	
TTG GTT TTG TTT TCG TTT TTT AAC ATT GCA TTA TGT GCT CCA AGG AAA	163
Leu Val Leu Phe Ser Phe Phe Asn Ile Ala Leu Cys Ala Pro Arg Lys	
-10 -5 1	
MMM CON ACT AND	
TTT GCA AGA AAG Phe Ala Arg Lys	175
5	
·	
(2) INFORMATION FOR SEQ ID NO: 163:	
/:) CEOURNER CHARACTERICO.	
(i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 383 base pairs	
(B) TYPE: NUCLEIC ACID	
(C) STRANDEDNESS: DOUBLE	
(D) TOPOLOGY: LINEAR	
(ii) MOLECULE TYPE: CDNA	
(vi) ORIGINAL SOURCE:	
(A) ORGANISM: Homo Sapiens	
(F) TISSUE TYPE: Pancreas	
(ix) FEATURE:	
(A) NAME/KEY: other	
(B) LOCATION: 49192	
(C) IDENTIFICATION METHOD: blastn	
(D) OTHER INFORMATION: identity 98	
region 43186	
id R38459	
est	
(ix) FEATURE:	
(1x) FEATURE: (A) NAME/KEY: other	
(B) LOCATION: 188240	
···	

137

(C) IDENTIFICATION METHOD: blastn (D) OTHER INFORMATION: identity 100 region 183235 id R38459 est																
	(12	(2 (1	3) L(C) II	AME/I OCATI DENTI	ON: [FIC]	othe 7! ATION	50 N ME:	re	denti egior 1 R38	ity] n 2						
	(ix) FEATURE: (A) NAME/KEY: sig_peptide (B) LOCATION: 60116 (C) IDENTIFICATION METHOD: Von Heijne matrix (D) OTHER INFORMATION: score 6.8 seq AIVALAVCAALHA/SE (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 163: AATTTCCTGA TCGAACAGCC TCACTTGTGT TGCTGTCAGT GCCAGTAGGG CAGGCAGGA 59															
AATT	TCC1	rga 1	rcga.	ACAGO	CC TO	CACT	rgtg'	r TGO	CTGTO	CAGT	GCC	AGTA	GGG (CAGGO	CAGGA	59
								GTG Val								107
CTA Leu	CAT His	GCC Ala	TCA Ser 1	GAA Glu	GCC Ala	ATA Ile	CTT Leu 5	CCC Pro	ATT Ile	GCC Ala	TCC Ser	AGC Ser 10	TGT Cys	TGC Cys	ACG Thr	155
								AGG Arg								203
TGT Cys 30	CGC Arg	ATC Ile	CAG Gln	AGA Arg	GCT Ala 35	GAT Asp	GGG Gly	GAT Asp	TGT Cys	GAC Asp 40	TTG Leu	GCT Ala	GCT Ala	GTC Val	ATC Ile 45	251
								TGT Cys								299
								GCT Ala 70								347
								CAT His								383
(2)	INF	ORMA	TION	FOR	SEQ	ID:	NO:	164:								

(2) IN

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 352 base pairs

138

- (B) TYPE: NUCLEIC ACID(C) STRANDEDNESS: DOUBLE(D) TOPOLOGY: LINEAR
- (ii) MOLECULE TYPE: CDNA
- (vi) ORIGINAL SOURCE:
 - (A) ORGANISM: Homo Sapiens
 - (F) TISSUE TYPE: Lung
- (ix) FEATURE:
 - (A) NAME/KEY: other
 - (B) LOCATION: 2..354
 - (C) IDENTIFICATION METHOD: blastn
 - (D) OTHER INFORMATION: identity 99

region 2..354 id HUM517F10B

est

- (ix) FEATURE:
 - (A) NAME/KEY: other
 - (B) LOCATION: 75..230
 - (C) IDENTIFICATION METHOD: blastn
 - (D) OTHER INFORMATION: identity 100

region 1..156

id H04128

.

- (ix) FEATURE:
 - (A) NAME/KEY: other
 - (B) LOCATION: 225..343
 - (C) IDENTIFICATION METHOD: blastn
 - (D) OTHER INFORMATION: identity 98

region 152..270

id H04128

est

- (ix) FEATURE:
 - (A) NAME/KEY: other
 - (B) LOCATION: 309..354
 - (C) IDENTIFICATION METHOD: blastn
 - (D) OTHER INFORMATION: identity 97

region 1..46 id AA099288

est

- (ix) FEATURE:
 - (A) NAME/KEY: sig_peptide
 - (B) LOCATION: 29..88
 - (C) IDENTIFICATION METHOD: Von Heijne matrix
 - (D) OTHER INFORMATION: score 6.8

seq GLLWMLFVSELRA/AT

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 164:

AAVCAGTTGG AGCTGGTGCA CAGGAAGG ATG AGG AAG ACC AGG CTC TGG GGG

Met Arg Lys Thr Arg Leu Trp Gly

-20 -15

CTG CTG TGG ATG CTC TTT GTC TCA GAA CTC CGA GCT GCA ACT AAA TTA 100

Leu	Leu	Trp -10	Met	Leu	Phe	Val	Ser -5	Glu	Leu	Arg	Ala	Ala 1	Thr	Lys	Leu	
				TAT Tyr												148
				CTA Leu 25												196
				GGA Gly												244
				TCC Ser												292
				CAT His												340
	VAA Xaa															352

(2) INFORMATION FOR SEQ ID NO: 165:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 293 base pairs

(B) TYPE: NUCLEIC ACID

(C) STRANDEDNESS: DOUBLE

(D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: CDNA

(vi) ORIGINAL SOURCE:

(A) ORGANISM: Homo Sapiens(F) TISSUE TYPE: Thyroid

(ix) FEATURE:

(A) NAME/KEY: other

(B) LOCATION: 36..241

(C) IDENTIFICATION METHOD: blastn

(D) OTHER INFORMATION: identity 98

region 1..206 id T08694

est

(ix) FEATURE:

(A) NAME/KEY: other

(B) LOCATION: 33..233

(C) IDENTIFICATION METHOD: blastn

(D) OTHER INFORMATION: identity 98

region 1..201 id AA095665

est

	(ix) FE	ATUE	RE:												
	•	. (A) NA	ME/E	ŒY:	othe	er									
						56										
								HOD:	h)a	etn						
						RMAT				ity 9	7					
		()	, 01	шык	TIVE) ICLINI	LON									
										18.	. 20.	3				
									1 R35	300						
								es	st.							
				_												
	(1X	:) FE														
		•	•			othe										
						97										
		-	-					CHOD:								
		(D	ro (e	HER	INFO	RMAI	'ION:	: ic	lenti	ity 9	8					
								re	egion	1 l	115					
								ic	i AAC	9303	19					
								es	it							
	(ix) FE	ATUF	RE:												
		(A	.) NA	ME/F	ŒY:	othe	er									
		(B) LC	CATI	ON:	118.	.292	2	•							
		•						CHOD:								
		(D	ro (HER	INFO	RMAT	'ION:	ic	lenti	ity S	5					
								re	gior	ı İ.,	175					
								ic	HSC	CIGAI	.21					
								es	st							
	(ix	:) FE														
		(A) NA	ME/F	ŒY:	sig_	pept	ide								
						102.										
		(C) IC	ENT	FICE	MOITA	ME1	HOD:	Vor	n Hei	.jne	matr	ix			
		(D) OI	HER	INFO	RMAT	CION	: sc	core	6.8	_					
								se	eq VS	SLVLI	MPG!	PCDG/	LF			
	(xi) SE	QUE!	ICE I	DESCE	RIPTI	ON:	SEQ	ID N	NO: 1	65:					
															•	
CTC	CTTC	TG C	CTG	ATTO	G GA	AAGCO	CTG	CAAC	GAC	AACC	GGC'	rggg	STC (CTTG	CGCGCC	60
GCGC	SCTCA	GG G	AGG	AGCA	CC GA	CTGC	CGCCC	G CAC	CCCT	GAGA	G A	rg gr	T G	GT GO	CC ATG	116
											Me	et Va	al G	ly Al	la Met	
													- 3	20		
TGG	AAG	GTG	ATT	GTT	TCG	CTG	GTC	CTG	TTG	ATG	CCT	GGC	CCC	TGT	GAT	164
Trp	Lys	Val	Ile	Val	Ser	Leu	Val	Leu	Leu	Met	Pro	Gly	Pro	Cys	Asp	
		-15					-10					-5		•		
GGG	CTG	TTT	CGC	TCC	CTA	TAC	AGA	AGT	GTT	BBC	ATG	CCA	CCT	AAG	GGA	212
Gly	Leu	Phe	Arg	Ser	Leu	Tyr	Arg	Ser	Val	Xaa	Met	Pro	Pro	Lvs	Glv	
•	1		•		5	•	•			10				,	15	
GAC	TCA	GGA	CAG	CCA	TTA	TTT	CTC	ACC	CCT	TAC	ATT	GAA	GCT	GGG	AAG	260
Asp	Ser	Glv	Gln	Pro	Leu	Phe	Leu	Thr	Pro	Tvr	Tle	Glu	Ala	Glv	Luc	200
				20					25	-] -		014	****	30	ByS	
														50		
ATC	CAA	AAA	GGN	AGA	GAA	YNN	RGT	TTG	GTC	GGG						293
	Gln															ر و ے
			35					40		1						

141

(2) INFORMATION FOR SEQ ID NO: 166:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 263 base pairs
 - (B) TYPE: NUCLEIC ACID
 - (C) STRANDEDNESS: DOUBLE
 - (D) TOPOLOGY: LINEAR
- (ii) MOLECULE TYPE: CDNA
- (vi) ORIGINAL SOURCE:
 - (A) ORGANISM: Homo Sapiens
 - (F) TISSUE TYPE: Lung (cells)
- (ix) FEATURE:
 - (A) NAME/KEY: other
 - (B) LOCATION: complement(1..158)
 - (C) IDENTIFICATION METHOD: blastn
 - (D) OTHER INFORMATION: identity 98

region 160..317 id N66156

est

- (ix) FEATURE:
 - (A) NAME/KEY: other
 - (B) LOCATION: complement (155..262)
 - (C) IDENTIFICATION METHOD: blastn
 - (D) OTHER INFORMATION: identity 99

region 55..162 id N66156

est

- (ix) FEATURE:
 - (A) NAME/KEY: other
 - (B) LOCATION: 167..262
 - (C) IDENTIFICATION METHOD: blastn
 - (D) OTHER INFORMATION: identity 96

region 8..103

id HUMGS02822

est

- (ix) FEATURE:
 - (A) NAME/KEY: sig_peptide
 - (B) LOCATION: 81..200
 - (C) IDENTIFICATION METHOD: Von Heijne matrix
 - (D) OTHER INFORMATION: score 6.1

seq ICIGILVLPFIRC/CC

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 166:

AAGAATAGGA TNATTCACTG GTAATAATAG AGTCATTAAG AAATATTAAG CATTGCAGCT 60

AAAAATTGAA CAACCCTGTA ATG ATA CAT TTA AGA ATT ATT CAA AGG TGC TAC 113
Met Ile His Leu Arg Ile Ile Gln Arg Cys Tyr

-40 -35 -30

ATG GCA GGG TTA GAG AAT AAA AAG AAC GTG GTG TTT GAA GCA AAA CAG
Met Ala Gly Leu Glu Asn Lys Lys Asn Val Val Phe Glu Ala Lys Gln

-25 -20 -15

ATC TGT ATT GGC ATC TTG GTT CTC CCT TTT ATC AGA TGT TGT TGC CTT 209 Ile Cys Ile Gly Ile Leu Val Leu Pro Phe Ile Arg Cys Cys Leu -10 -5 GTG CAA ATC ACA TTT TCT CTG AGT CTC CAT TTT CTC ATT TAT AAC ATG 257 Val Gln Ile Thr Phe Ser Leu Ser Leu His Phe Leu Ile Tyr Asn Met CGG CGG 263 Arg Arg 20 (2) INFORMATION FOR SEQ ID NO: 167: (i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 347 base pairs (B) TYPE: NUCLEIC ACID (C) STRANDEDNESS: DOUBLE (D) TOPOLOGY: LINEAR (ii) MOLECULE TYPE: CDNA (vi) ORIGINAL SOURCE: (A) ORGANISM: Homo Sapiens (F) TISSUE TYPE: Lung (cells) (ix) FEATURE: (A) NAME/KEY: other (B) LOCATION: 61..344 (C) IDENTIFICATION METHOD: blastn (D) OTHER INFORMATION: identity 98 region 18..301 id H73135 est (ix) FEATURE: (A) NAME/KEY: other (B) LOCATION: 49..109 (C) IDENTIFICATION METHOD: blastn (D) OTHER INFORMATION: identity 100 region 1..61 id AA251602 (ix) FEATURE: (A) NAME/KEY: sig_peptide (B) LOCATION: 54..119 (C) IDENTIFICATION METHOD: Von Heijne matrix (D) OTHER INFORMATION: score 5.9 seq LIYILWQLTGSAA/SG (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 167: AATACCTAAG AGGGAAGTGG CTTCATTTCA GTGGCTGACT TCCAGAGAGC AAT ATG GCT GGT TCC CCA ACA TGC CTC ACC CTC ATC TAT ATC CTT TGG CAG CTC 104 Ala Gly Ser Pro Thr Cys Leu Thr Leu Ile Tyr Ile Leu Trp Gln Leu

143

•										
	-20			-15			-10			
			GCC Ala							152
			ACT Thr							200
			ACC Thr							248
			ATC Ile							296
			GGA Gly							344
GGG Gly										347

(2) INFORMATION FOR SEQ ID NO: 168:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 394 base pairs

(B) TYPE: NUCLEIC ACID

(C) STRANDEDNESS: DOUBLE

(D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: CDNA

(vi) ORIGINAL SOURCE:

(A) ORGANISM: Homo Sapiens

(F) TISSUE TYPE: Thyroid

(ix) FEATURE:

(A) NAME/KEY: other

(B) LOCATION: 86..206

(C) IDENTIFICATION METHOD: blastn

(D) OTHER INFORMATION: identity 100

region 81..201 id AA159880

est

(ix) FEATURE:

(A) NAME/KEY: other

(B) LOCATION: 5..85

(C) IDENTIFICATION METHOD: blastn

(D) OTHER INFORMATION: identity 100

region 1..81 id AA159880

est

(ix) FEATURE:

(A) NAME/KEY: other

144

(B) LOCATION: 244..309

(C) IDENTIFICATION METHOD: blastn

(D) OTHER INFORMATION: identity 100

region 239..304 id AA159880

est

(ix) FEATURE:

(A) NAME/KEY: other

(B) LOCATION: 239..354

(C) IDENTIFICATION METHOD: blastn

(D) OTHER INFORMATION: identity 94

region 1..116 id AA109004

est

(ix) FEATURE:

(A) NAME/KEY: sig_peptide

(B) LOCATION: 56..232

(C) IDENTIFICATION METHOD: Von Heijne matrix

(D) OTHER INFORMATION: score 5.7

seq ALLDLCAAPXGWL/QV

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 168:

AAGTGTGGAG AAAGCGGCTC TGGGTCTAGA TTGAGGGATA CTCCCCCTTT CCACC ATG 58 GGC AAG AAG GGC AAA GTT GGC AAG AGC CGA CGA GAC AAG TTT TAT CAC Gly Lys Lys Gly Lys Val Gly Lys Ser Arg Arg Asp Lys Phe Tyr His TTG GCG AAG GAG ACG GGT TAC CGT TCC CGA TCT GCT TTC AAG CTG ATC Leu Ala Lys Glu Thr Gly Tyr Arg Ser Arg Ser Ala Phe Lys Leu Ile -35 CAG CTC AAT CGC CGC TTT CAG TTC CTG CAG AAA GCC CGA GCC TTG CTG 202 Gln Leu Asn Arg Arg Phe Gln Phe Leu Gln Lys Ala Arg Ala Leu Leu -20 GAC CTG TGT GCT GCG CCA SGG GGA TGG CTG CAG GTA GCT GCC AAG TTT 250 Asp Leu Cys Ala Ala Pro Xaa Gly Trp Leu Gln Val Ala Ala Lys Phe ATG CCT GTA TCC AGC CTT ATT GTG GGA GTG GAC CTG GTT CCA ATC AAG 298 Met Pro Val Ser Ser Leu Ile Val Gly Val Asp Leu Val Pro Ile Lys CCT CTC CCC AAT GTG GTG ACT CTC CAG GAG GAC ATC ACA ACA GAA CGT 346 Pro Leu Pro Asn Val Val Thr Leu Gln Glu Asp Ile Thr Thr Glu Arg 30 TGT ARG CAA AGG CAC TGG ACA TCA GCC TCA GCT CTG GAR AGG AAG ATG 394 Cys Xaa Gln Arg His Trp Thr Ser Ala Ser Ala Leu Glu Arg Lys Met 40 45

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 299 base pairs

(B) TYPE: NUCLEIC ACID

(C) STRANDEDNESS: DOUBLE

(D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: CDNA

(vi) ORIGINAL SOURCE:

(A) ORGANISM: Homo Sapiens

(D) DEVELOPMENTAL STAGE: Fetal

(F) TISSUE TYPE: liver

(ix) FEATURE:

(A) NAME/KEY: other

(B) LOCATION: 22..296

(C) IDENTIFICATION METHOD: blastn

(D) OTHER INFORMATION: identity 96

region 68..342 id R21563

est

(ix) FEATURE:

(A) NAME/KEY: other

(B) LOCATION: 23..176

(C) IDENTIFICATION METHOD: blastn

(D) OTHER INFORMATION: identity 95

region 1..154 id R23529

est

(ix) FEATURE:

(A) NAME/KEY: other

(B) LOCATION: 158..296

(C) IDENTIFICATION METHOD: blastn

(D) OTHER INFORMATION: identity 94

region 135..273

id R23529

est

(ix) FEATURE:

(A) NAME/KEY: other

(B) LOCATION: 3..271

(C) IDENTIFICATION METHOD: blastn

(D) OTHER INFORMATION: identity 98

region 1..269 id HSC2LG091

est

(ix) FEATURE:

(A) NAME/KEY: other

(B) LOCATION: 22..179

(C) IDENTIFICATION METHOD: blastn

(D) OTHER INFORMATION: identity 94

region 68..225

id R20326

est

(ix) FEATURE:

(A) NAME/KEY: other

(B) LOCATION: 180..262

(C) IDENTIFICATION METH (D) OTHER INFORMATION:	
(ix) FEATURE: (A) NAME/KEY: other (B) LOCATION: 198296 (C) IDENTIFICATION METH (D) OTHER INFORMATION:	. – –
(ix) FEATURE:(A) NAME/KEY: other(B) LOCATION: 2678(C) IDENTIFICATION METH(D) OTHER INFORMATION:	
<pre>(ix) FEATURE: (A) NAME/KEY: other (B) LOCATION: 78108 (C) IDENTIFICATION METH (D) OTHER INFORMATION:</pre>	
<pre>(ix) FEATURE: (A) NAME/KEY: other (B) LOCATION: 115145 (C) IDENTIFICATION METH (D) OTHER INFORMATION:</pre>	
<pre>(ix) FEATURE: (A) NAME/KEY: sig_pepti (B) LOCATION: 75170 (C) IDENTIFICATION METH (D) OTHER INFORMATION:</pre>	OD: Von Heijne matrix
(xi) SEQUENCE DESCRIPTION: S	
AACGGAGGGC CAGAGAGTCA CGGCGGTTTT	
GTTTCCCTTT CAAG ATG CCA CTT TCA G Met Pro Leu Ser A -30	SAC TTT ATT CTG GCT CTK AAG GAC 110 Asp Phe Ile Leu Ala Leu Lys Asp -25
AAT CCC TAC TTT GGG GCT GGA TTT G Asn Pro Tyr Phe Gly Ala Gly Phe G -20 -15	GGG CTG GTG WGT GTG GGC ACA GCC 158 Gly Leu Val Xaa Val Gly Thr Ala -10 -5

.

CTG Leu	GCC Ala	CTG Leu	GCC Ala	Arg	AAG Lys	GGT Gly	GTC Val	Gln	CTG Leu	GGC Gly	CTG Leu	GTG Val	Ala	TTC Phe	CGG Arg	206
CGC Arg	CAT His	TAC Tyr 15	ATG Met	ATC Ile	ACA Thr	CTG Leu	GAA Glu 20	5 GTC Val	CCT Pro	GCT Ala	CGA Arg	GAC Asp 25	10 AGG Arg	AGC Ser	TAT Tyr	254
GSC Xaa	TGG Trp 30	TTG Leu	CTT Leu	AGC Ser	TGG Trp	CTC Leu 35	ACC Thr	CGC Arg	CAC His	AGT Ser	ACC Thr 40	CGT Arg	ACT Thr	GGG Gly		299
(2)	(ii) (i)	SEQ (1 (1 (1 (1 (1 (1 (1 (1 (1 (1 (1 (1 (1	QUENCA) LI C) ST C) ST C) CLECT RIGINA OLECT RIGINA OLECT RIGINA OLECT CATURA OLECT CATURA OLECT CATURA OLECT OL	CE CHENGTH (PE: PRANT POPOLO JLE T NAL S RGANT ISSUE RE: AME/E DENT I THER	HARACHE 35 NUCIDEDNE DGY: CYPE: COURCESM: CYPE: CON: INFO	CTERNS S9 ba LEIC CSS: LINE CDN CE: Homo PE: I OTHE 321. ATION DRMA	DOUBLEAR NA DO Sarge 1.358 N MET FION:	CS: Dairs Diens SILE Diens SIHOD: CONTRACT CONTR	: bladenti	astn ity 9 n 241 ClD10 n Hei 5.6 /FLV	l27) ijne	mat				
ACC:	CTG	GGA (GAAGI	MRTC	CC AC	GCCC	CAGAI	A TTO	CCA	GGAG	TCT	CCGT	rcg (STGA'	rcagca	60
CTG	AACA	CCG 1	AGGA	CTCAG					ly L					ne L	TC GTT eu Val 10	113
GCT Ala	ATT Ile	ATA Ile	AAA Lys -5	GGT Gly	GTC Val	CAG Gln	TGT Cys	CAG Gln l	GCA Ala	CAG Gln	CTC Leu	GAG Glu 5	GAA Glu	TCT Ser	GGG Gly	161
	GGC Gly 10															209

		CTC Leu						257
		MTG Xaa 45						305
		GGG Gly						353
GAC Asp								359

(2) INFORMATION FOR SEQ ID NO: 171:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 177 base pairs
 - (B) TYPE: NUCLEIC ACID
 - (C) STRANDEDNESS: DOUBLE
 - (D) TOPOLOGY: LINEAR
- (ii) MOLECULE TYPE: CDNA
- (vi) ORIGINAL SOURCE:
 - (A) ORGANISM: Homo Sapiens
 - (F) TISSUE TYPE: Lung (cells)
- (ix) FEATURE:
 - (A) NAME/KEY: other
 - (B) LOCATION: 145..176
 - (C) IDENTIFICATION METHOD: blastn
 - (D) OTHER INFORMATION: identity 90

region 56..87 id N87931

est

- (ix) FEATURE:
 - (A) NAME/KEY: other
 - (B) LOCATION: 145..176
 - (C) IDENTIFICATION METHOD: blastn
 - (D) OTHER INFORMATION: identity 90

region 287..318

id C17084

est

- (ix) FEATURE:
 - (A) NAME/KEY: sig_peptide
 - (B) LOCATION: 46..108
 - (C) IDENTIFICATION METHOD: Von Heijne matrix
 - (D) OTHER INFORMATION: score 5.3

seq SLFSSLPIFLTWA/HI

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 171:

149

Met Ile Leu Arg -20 AAA AGA TCA TGC TCA TTA TTT TCT TCC CTC CCA ATC TTT CTA ACA TGG 105 Lys Arg Ser Cys Ser Leu Phe Ser Ser Leu Pro Ile Phe Leu Thr Trp -10 GCC CAC ATA AAA CGT GTC CCC CTT CTG CMA ACA TCC CTT CAC ACC GCC 153 Ala His Ile Lys Arg Val Pro Leu Leu Xaa Thr Ser Leu His Thr Ala 5 CAC AAC GGC CAC CCC CAC TAC GGG 177 His Asn Gly His Pro His Tyr Gly 20

- (2) INFORMATION FOR SEQ ID NO: 172:
 - (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 188 base pairs
 - (B) TYPE: NUCLEIC ACID
 - (C) STRANDEDNESS: DOUBLE
 - (D) TOPOLOGY: LINEAR
 - (ii) MOLECULE TYPE: CDNA
 - (vi) ORIGINAL SOURCE:
 - (A) ORGANISM: Homo Sapiens
 - (F) TISSUE TYPE: Colon
 - (ix) FEATURE:
 - (A) NAME/KEY: other
 - (B) LOCATION: 21..137
 - (C) IDENTIFICATION METHOD: blastn
 - (D) OTHER INFORMATION: identity 96 region 178..294 id AA148442

est

- (ix) FEATURE:
 - (A) NAME/KEY: other
 - (B) LOCATION: 144..184
 - (C) IDENTIFICATION METHOD: blastn
 - (D) OTHER INFORMATION: identity 97

region 303..343 id AA148442

est

- (ix) FEATURE:
 - (A) NAME/KEY: sig_peptide
 - (B) LOCATION: 99..146
 - (C) IDENTIFICATION METHOD: Von Heijne matrix
 - (D) OTHER INFORMATION: score 5

seq GLMFVKLVNPCSG/EG

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 172:

150

TCCTGGTTTC AGAATATTTA AAAGATGCTT CAAAGAAG ATG AAA AAT GGG CTA ATG 116

Met Lys Asn Gly Leu Met
-15

TTT GTA AAA CTG GTT AAC CCC TGT TCA GGA GAA GGA GCC ATT TAC TTG

Phe Val Lys Leu Val Asn Pro Cys Ser Gly Glu Gly Ala Ile Tyr Leu

-10

-5

1

5

TTC AAT ATG TGT CTA CAG CAG CGG Phe Asn Met Cys Leu Gln Gln Arg 10

188

(2) INFORMATION FOR SEQ ID NO: 173:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 182 base pairs
- (B) TYPE: NUCLEIC ACID
- (C) STRANDEDNESS: DOUBLE
- (D) TOPOLOGY: LINEAR
- (ii) MOLECULE TYPE: CDNA
- (vi) ORIGINAL SOURCE:
 - (A) ORGANISM: Homo Sapiens
 - (F) TISSUE TYPE: Lung (cells)
- (ix) FEATURE:
 - (A) NAME/KEY: other
 - (B) LOCATION: 37..181
 - (C) IDENTIFICATION METHOD: blastn
 - (D) OTHER INFORMATION: identity 95 region 17..161 id W24468

est

- (ix) FEATURE:
 - (A) NAME/KEY: other
 - (B) LOCATION: 62..181
 - (C) IDENTIFICATION METHOD: blastn
 - (D) OTHER INFORMATION: identity 100 region 3..122 id W38688 est

(ix) FEATURE:

- (A) NAME/KEY: other
- (B) LOCATION: 62..181
- (C) IDENTIFICATION METHOD: blastn
- (D) OTHER INFORMATION: identity 97 region 3..122 id W80906

est

- (ix) FEATURE:
 - (A) NAME/KEY: other
 - (B) LOCATION: 71..181
 - (C) IDENTIFICATION METHOD: blastn
 - (D) OTHER INFORMATION: identity 100 region 1..111

151

id AA213429 est

•	iх	١	CC	יתי	וזיד	RE.

(A) NAME/KEY: other

(B) LOCATION: 71..181

(C) IDENTIFICATION METHOD: blastn

(D) OTHER INFORMATION: identity 100

region 3..113 id AA054464

est

(ix) FEATURE:

(A) NAME/KEY: sig_peptide

- (B) LOCATION: 66..110
- (C) IDENTIFICATION METHOD: Von Heijne matrix
- (D) OTHER INFORMATION: score 4.9

seq AVVFVFSLLDCCA/LI

182

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 173:

TTCAGGGGTG GGTCGGGGCA TCCGAGCGGG TTTGACGGAA GGAGCGGCGG CGACGGAGGA 60

GGAGG ATG GAG GCG GTG GTG TTC GTC TTC TCT CTC CTC GAT TGT TGC GCG 110

Met Glu Ala Val Val Phe Val Phe Ser Leu Leu Asp Cys Cys Ala

-15

-10

-5

CTC ATC TTC CTC TCG GTC TAC TTC ATA ATT ACA TTG TCT GAT TTA GAA

Leu Ile Phe Leu Ser Val Tyr Phe Ile Ile Thr Leu Ser Asp Leu Glu

1 5 10 158

TGT GAT TAC ATT AAT GCT AGA TCG
Cys Asp Tyr Ile Asn Ala Arg Ser
20

(2) INFORMATION FOR SEQ ID NO: 174:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 349 base pairs
 - (B) TYPE: NUCLEIC ACID
 - (C) STRANDEDNESS: DOUBLE
 - (D) TOPOLOGY: LINEAR
- (ii) MOLECULE TYPE: CDNA
- (vi) ORIGINAL SOURCE:
 - (A) ORGANISM: Homo Sapiens
 - (F) TISSUE TYPE: Pancreas

(ix) FEATURE:

- (A) NAME/KEY: other
- (B) LOCATION: 141..310
- (C) IDENTIFICATION METHOD: blastn ·
- (D) OTHER INFORMATION: identity 100 region 118..287 id W25106

est

PCT/IB98/01233 WO 99/06439

(ix) FEATURE:

(A) NAME/KEY: other (B) LOCATION: 22..105

(C) IDENTIFICATION METHOD: blastn

(D) OTHER INFORMATION: identity 100

region 1..84 id W25106 est

152

(ix) FEATURE:

(A) NAME/KEY: other

(B) LOCATION: 107..136

(C) IDENTIFICATION METHOD: blastn

(D) OTHER INFORMATION: identity 100

region 85..114 id W25106

est

(ix) FEATURE:

(A) NAME/KEY: other

(B) LOCATION: 142..351

(C) IDENTIFICATION METHOD: blastn

(D) OTHER INFORMATION: identity 99

region 121..330 id N31560

est

(ix) FEATURE:

(A) NAME/KEY: other

(B) LOCATION: 24..112

(C) IDENTIFICATION METHOD: blastn

(D) OTHER INFORMATION: identity 96

region 1..89 id N31560

est

(ix) FEATURE:

(A) NAME/KEY: other

(B) LOCATION: 36..351

(C) IDENTIFICATION METHOD: blastn

(D) OTHER INFORMATION: identity 98

region 1..316 id W17036

est

(ix) FEATURE:

(A) NAME/KEY: other (B) LOCATION: 141..351

(C) IDENTIFICATION METHOD: blastn

(D) OTHER INFORMATION: identity 99

region 105..315 id AA039274

est

(ix) FEATURE:

(A) NAME/KEY: other (B) LOCATION: 37..112

(C) IDENTIFICATION METHOD: blastn

(D) OTHER INFORMATION: identity 97

region 1..76 id AA039274

est

153

		·				
(ix)	FEATURE: (A) NAME/KEY: (B) LOCATION: (C) IDENTIFICAT (D) OTHER INFOR	complement (FION METHOD RMATION: i r	: blastn			
(ix)	FEATURE: (A) NAME/KEY: C (B) LOCATION: C (C) IDENTIFICAT (D) OTHER INFOR	complement (TION METHOD RMATION: i r	: blastn			
(ix)	FEATURE: (A) NAME/KEY: C (B) LOCATION: C (C) IDENTIFICAT (D) OTHER INFOR	complement (FION METHOD RMATION: i r	: blastn			
	FEATURE: (A) NAME/KEY: s (B) LOCATION: 1 (C) IDENTIFICAT (D) OTHER INFOR	L64331 FION METHOD RMATION: s	core 4.7 eq FACVPGA	ASXTTLA/FP		
	SEQUENCE DESCRI					
	A AGGGGCGGGA CC1					60
	C GCTGCCGTCG CCA					120
AATATGAAA	A AGCAGAGCGA CTO	CGGTTAAG GG	AAAGCGCC G		GGC TTT Gly Phe	175
Leu Leu Pr	CG CCC GCA AGC A ro Pro Ala Ser A 50					223
	GG CAA ACG AGA A		Asn Pro S			271
	CA ACC CTC TTG (ro Thr Leu Leu I -15					319

349

ACC ACG CTC GCG TTT CCT CCT GTA GTG CTC

Thr Thr Leu Ala Phe Pro Pro Val Val Leu

PCT/IB98/01233

5

1

(2) INFORMATION FOR SEQ ID NO: 175:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 273 base pairs
- (B) TYPE: NUCLEIC ACID
- (C) STRANDEDNESS: DOUBLE
- (D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: CDNA

(vi) ORIGINAL SOURCE:

- (A) ORGANISM: Homo Sapiens
- (F) TISSUE TYPE: Colon

(ix) FEATURE:

- (A) NAME/KEY: other
- (B) LOCATION: 42..271
- (C) IDENTIFICATION METHOD: blastn
- (D) OTHER INFORMATION: identity 96

region 1..230 id R20112

est

(ix) FEATURE:

- (A) NAME/KEY: other
- (B) LOCATION: 79..271
- (C) IDENTIFICATION METHOD: blastn
- (D) OTHER INFORMATION: identity 92

region 91..283

id W35748

est

(ix) FEATURE:

- (A) NAME/KEY: other
- (B) LOCATION: 108..222
- (C) IDENTIFICATION METHOD: blastn
- (D) OTHER INFORMATION: identity 93

region 78..192 id AA074652

est

(ix) FEATURE:

- (A) NAME/KEY: other
- (B) LOCATION: 31..81
- (C) IDENTIFICATION METHOD: blastn
- (D) OTHER INFORMATION: identity 100

region 1..51

id AA074652

est

(ix) FEATURE:

- (A) NAME/KEY: other
- (B) LOCATION: 222..271
- (C) IDENTIFICATION METHOD: blastn
- (D) OTHER INFORMATION: identity 100

region 191..240

id AA074652

est

	(i)	(E	A) NA B) LO C) II	AME/H DCÁTI DENTI	(EY: ION: IFICA INF(79 TION	.271 N ME1	: io	denti egior d W35	ty 9	91 283	3				
	(i)	(E	A) NA B) LO C) II	AME/F CATI CENTI	KEY: ION: IFICA INFO	79 TION	271 1 MET	i i	denti egior d W74	ty 9	92 285	;				
		1) () ()	A) NA B) LO C) II D) OT	AME/F DCATI DENTI THER	(EY: (ON: (FIC) (INFO	94. ATION DRMAT	150 MET TION:	CHOD: se	core eq SV	4.5 /PLL1	rdaat					
AAAC	SAGTO	STG 7	CTGC	CGGG	AG AA	\AGA(GAG	A ATO	CGCCC	CAAG	CGGC	CTC	GA A	AGTCC	CAGGG	60
AGTO	GGAGC	SCC (CCCG	CCGT	G AC	GCCG1	rgtgo	G TG1						Met	G TCT Ser	114
GTG Val	CCC Pro	CTG Leu -10	CTC Leu	ACC Thr	GAT Asp	GCT Ala	GCC Ala -5	ACC Thr	GTG Val	TCT Ser	GGA Gly	GCT Ala 1	GAG Glu	CGG Arg	GAA Glu	162
ACG Thr 5	GCC Ala	GCG Ala	GTT Val	ATT Ile	TTT Phe 10	TTA Leu	CAT His	GGA Gly	CTT Leu	GGA Gly 15	GAC Asp	ACA Thr	GGG Gly	CAC His	AGC Ser 20	210
TGG Trp	GCT Ala	GAC Asp	GCC Ala	CTC Leu 25	TCC Ser	ACC Thr	ATC Ile	CGG Arg	CTC Leu 30	CCT Pro	CAC His	GTC Val	AAG Lys	TAC Tyr 35	ATC Ile	258
	CCC Pro															273

(2) INFORMATION FOR SEQ ID NO: 176:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 371 base pairs
 - (B) TYPE: NUCLEIC ACID
 - (C) STRANDEDNESS: DOUBLE
 (D) TOPOLOGY: LINEAR

```
(ii) MOLECULE TYPE: CDNA
```

- (vi) ORIGINAL SOURCE:
 - (A) ORGANISM: Homo Sapiens
 - (F) TISSUE TYPE: Large intestine
- (ix) FEATURE:
 - (A) NAME/KEY: other
 - (B) LOCATION: complement (53..194)
 - (C) IDENTIFICATION METHOD: blastn
 - (D) OTHER INFORMATION: identity 98

region 178..319

id R00081

est

- (ix) FEATURE:
 - (A) NAME/KEY: other
 - (B) LOCATION: complement(202..322)
 - (C) IDENTIFICATION METHOD: blastn
 - (D) OTHER INFORMATION: identity 98

region 52..172

id R00081

est

- (ix) FEATURE:
 - (A) NAME/KEY: other
 - (B) LOCATION: complement (322..372)
 - (C) IDENTIFICATION METHOD: blastn
 - (D) OTHER INFORMATION: identity 92

region 3..53

id R00081

est

- (ix) FEATURE:
 - (A) NAME/KEY: other
 - (B) LOCATION: complement(202..322)
 - (C) IDENTIFICATION METHOD: blastn
 - (D) OTHER INFORMATION: identity 98

region 76..196

id T53389

est

- (ix) FEATURE:
 - (A) NAME/KEY: other
 - (B) LOCATION: complement(104..194)
 - (C) IDENTIFICATION METHOD: blastn
 - (D) OTHER INFORMATION: identity 98

region 202..292

id T53389

est

- (ix) FEATURE:
 - (A) NAME/KEY: other
 - (B) LOCATION: complement(322..372)
 - (C) IDENTIFICATION METHOD: blastn

(D) OTHER INFORMATION: identity 92

region 27..77

id T53389

est

157

(ix) FEATURE:

(A) NAME/KEY: other

(B) LOCATION: complement(70..105)

(C) IDENTIFICATION METHOD: blastn

(D) OTHER INFORMATION: identity 100

region 292..327

id T53389

est

(ix) FEATURE:

(A) NAME/KEY: other

(B) LOCATION: complement(247..322)

(C) IDENTIFICATION METHOD: blastn

(D) OTHER INFORMATION: identity 98

region 70..145 id R50426

est

(ix) FEATURE:

(A) NAME/KEY: other

(B) LOCATION: complement(322..372)

(C) IDENTIFICATION METHOD: blastn

(D) OTHER INFORMATION: identity 92

region 21..71 id R50426

est

(ix) FEATURE:

(A) NAME/KEY: other

(B) LOCATION: complement(202..248)

(C) IDENTIFICATION METHOD: blastn

(D) OTHER INFORMATION: identity 97

region 145..191

id R50426

est

(ix) FEATURE:

(A) NAME/KEY: other

(B) LOCATION: 3..194

(C) IDENTIFICATION METHOD: blastn

(D) OTHER INFORMATION: identity 96

region 1..192

id H26655

est

(ix) FEATURE:

(A) NAME/KEY: other

(B) LOCATION: complement(225..322)

(C) IDENTIFICATION METHOD: blastn

(D) OTHER INFORMATION: identity 96

region 71..168

id R52030

est

(ix) FEATURE:

(A) NAME/KEY: other

(B) LOCATION: complement(322..372)

(C) IDENTIFICATION METHOD: blastn

(D) OTHER INFORMATION: identity 92

region 22..72

id R52030

est

			•											
	(ix)	(B) L(AME/K CATIO CENTI	ON: 18 FICAT:	ig_pept 8931 ION ME MATION	7 THOD: : sc	core		•					
	(xi)	SEQUE	NCE D	ESCRI	PTION:	SEQ	ID 1	10: 1	.76:					
CCG	rgagtc(G TACA	CTTGA	T GGC	rcctg	C TAC	STCC	GCCA	GAAG	GCAC	SAG (STCCA	\GGTGW	60
GGC	TGGAG	C CAATO	GGAA	G GTG	GCTGTG	A TTO	STCAC	CAA	TGAC	CATO	CT (GGAA	ACTGT	120
GTG	GGCCT	G TGGA	AACTT'	T GAC	GGGGAC	C AGA	ACCA	ATGA	TTGG	CATO	SAC 1	CCC	AGGAGA	180
AGC		ATG GWO							Ser					230
GCT Ala	GAT CA	AG TCA ln Ser	TCC Ser -25	ACC ACT Thr A	GG AAC rg Asn	GAA Glu	GAT Asp -20	TTC Phe	CTG Leu	AAG Lys	AAG Lys	ACC Thr -15	TGG Trp	278
		GG AGG rp Arg -10												326
		TA CCS eu Pro		Phe T										371
(2)	INFOR	MATION	FOR	SEQ I	D NO:	177:								

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 329 base pairs
 - (B) TYPE: NUCLEIC ACID
 - (C) STRANDEDNESS: DOUBLE
 - (D) TOPOLOGY: LINEAR
- (ii) MOLECULE TYPE: CDNA
- (vi) ORIGINAL SOURCE:
 - (A) ORGANISM: Homo Sapiens
 - (F) TISSUE TYPE: Pancreas
- (ix) FEATURE:
 - (A) NAME/KEY: other
 - (B) LOCATION: 238..329
 - (C) IDENTIFICATION METHOD: blastn
 - (D) OTHER INFORMATION: identity 90 region 18..109 id T29009

est

- (ix) FEATURE:
 - (A) NAME/KEY: other

99/0643	9					•	
_					159		
•	(C)		CATION I	ИЕТН	OD: blastn identity region 13 id R01547 est		
(ix)	(A) (B) (C)	NAME/KEY LOCATION IDENTIFE	: 238 CATION N	1 ЕТН	OD: blastn identity ! region 46 id T91432 est		
(ix)	(A) (B) (C) (D)	NAME/KEY LOCATION IDENTIFY OTHER IN	N: 249 CATION N NFORMATIO	290 ÆTH ON:	de OD: Von He: score 3.8 seq FLLLNO	CIVAVSQN/	
LAAA GG	CT(SATACTTC	CCAGGAC	GCG	GAGGTAACGG	GCCAGGGC	CA A
ACTTGG	ATI	rggtcgc	GTAGTTT	GG	CGGCCGGACC	TTAGAAAG	TC A
CTGTGC	: GGC	CCCTGCT	TCTGCGG	ATG	CTGAGGCACG	TAAAAAAA	TT ?
rcgcgg	CAT	TCTTGGC	CTGGCTT	CCT	GGCGTACASC	AAGTTCGG	AG (

AAAAAAAAG AAGCGACTTT 60 CGCTACTTG ACACATCTGC 120 GCGCCTGTG TGAAGAAGGG 180 AATTTCGCG GTGTTAACCG 240 CTGCTGTC ATG TTT CTT TTG CTA AAC TGC ATC GTC GCT GTG TCC CAA AAC 290 Met Phe Leu Leu Asn Cys Ile Val Ala Val Ser Gln Asn -10 ATG GGC ATC GGC AAG AAC GGG GAC CTG CCS VGG CCG CAG Met Gly Ile Gly Lys Asn Gly Asp Leu Pro Xaa Pro Gln

(2) INFORMATION FOR SEQ ID NO: 178:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 194 base pairs
 - (B) TYPE: NUCLEIC ACID
 - (C) STRANDEDNESS: DOUBLE
 - (D) TOPOLOGY: LINEAR
- (ii) MOLECULE TYPE: CDNA
- (vi) ORIGINAL SOURCE:
 - (A) ORGANISM: Homo Sapiens
 - (F) TISSUE TYPE: Lung
- (ix) FEATURE:
 - (A) NAME/KEY: other
 - (B) LOCATION: complement(163..192)

160

(C) IDENTIFICATION METHOD: blastn	
(D) OTHER INFORMATION: identity 90	
region 175204	
id H33662	
est	
(ix) FEATURE:	
(A) NAME/KEY: sig_peptide	
(B) LOCATION: 1556	
(C) IDENTIFICATION METHOD: Von Heijne matrix	
(D) OTHER INFORMATION: score 3.8	
seq LLVSAAPLGFGQG/VW	
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 178:	
AGGTTGTTTA CTGA ATG CTT CTG GTA TCT GCA GCC CCG CTG GGG TTC GGA	50
Met Leu Leu Val Ser Ala Ala Pro Leu Gly Phe Gly	
-10 -5	
CAC CCC CMC MCC AAM ACC COM MCA CAA CMA CAC CAC CAC CAC	
CAG GGG GTC TGG AAT AGG GCT TCA CAA CTA CAG CAG GGC TAS GAC CCT	98
Gln Gly Val Trp Asn Arg Ala Ser Gln Leu Gln Gln Gly Xaa Asp Pro	
1 5 10	
CMM CCC CCM CC3 3CC 3CC MCC 3C3 CCC CMC MCC 33C MCC MC	
CTT GGG GCT GGA AGG AGC TGG AGA GGC CTC TGC AAG CTG TCA CAG GCT	146
Leu Gly Ala Gly Arg Ser Trp Arg Gly Leu Cys Lys Leu Ser Gln Ala	
15 20 25 30	
CDD CCD CCD CCC NCD CCC DCN CCC DDD CCC DDD CNC NCN CNC NCN CCC	
CTT GGT GCT GGC ACT GGC TCA GGC TTT CAC ACA CAC ACA CGC GCA CCA	194
Leu Gly Ala Gly Thr Gly Ser Gly Phe His Thr His Thr Arg Ala Pro	
35 40 45	
(2) INFORMATION FOR SEQ ID NO: 179:	
(2) INTONESTION FOR SEQ 15 NO. 175.	
(i) SEQUENCE CHARACTERISTICS:	
(A) LENGTH: 187 base pairs	
(B) TYPE: NUCLEIC ACID	
(C) STRANDEDNESS: DOUBLE	
(D) TOPOLOGY: LINEAR	
(b) Totobodi. Bittant	
(ii) MOLECULE TYPE: CDNA	
(II) HODDOODD IIIB. ODNA	
(vi) ORIGINAL SOURCE:	
(A) ORGANISM: Homo Sapiens	
(D) DEVELOPMENTAL STAGE: Fetal	
(F) TISSUE TYPE: liver	
(I) IIOOOD IIIE. IIVEL	
(ix) FEATURE:	
(A) NAME/KEY: other	
(B) LOCATION: complement(2174)	
(C) IDENTIFICATION METHOD: blastn	
(D) OTHER INFORMATION: identity 100	
region 1173	
id M85423	
est	
esc	
(ix) FEATURE:	
(A) NAME/KEY: other	
(A) LOCATION: complement (01 192)	

161

(C) IDENTIFICATION METHOD: blastn
(D) OTHER INFORMATION: identity 98
region 467..559
id AA126476
est

(ix) FEATURE:

(A) NAME/KEY: other

(B) LOCATION: complement (28..89)

(C) IDENTIFICATION METHOD: blastn

(D) OTHER INFORMATION: identity 98

region 560..621 id AA126476

est

(ix) FEATURE:

(A) NAME/KEY: other

(B) LOCATION: complement(2..32)

(C) IDENTIFICATION METHOD: blastn

(D) OTHER INFORMATION: identity 90

region 616..646 id AA126476

est

(ix) FEATURE:

(A) NAME/KEY: other

(B) LOCATION: 47..183

(C) IDENTIFICATION METHOD: blastn

(D) OTHER INFORMATION: identity 100

region 1..137 id R33928 est

(ix) FEATURE:

(A) NAME/KEY: other

(B) LOCATION: 55..183

(C) IDENTIFICATION METHOD: blastn

(D) OTHER INFORMATION: identity 100

region 14..142 id H67425

est

(ix) FEATURE:

(A) NAME/KEY: other
(B) LOCATION: 56..183

(C) IDENTIFICATION METHOD: blastn

(D) OTHER INFORMATION: identity 100

region 1..128 id W04820

est

(ix) FEATURE:

(A) NAME/KEY: sig_peptide

(B) LOCATION: 80..127

(C) IDENTIFICATION METHOD: Von Heijne matrix

(D) OTHER INFORMATION: score 3.7

seq IALTLIPSMLSRA/AG

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 179:

WO 99/06439		PCT/IB98/01233
	162	•

W	O 99/06439		PCT/IB98/01233
	•	62	
ACC	TTCTTGT TATTTATGCT ATTCTCTTTG TGGCTC	CCATT CTTCTTTCAA TCTT	CTCAGC 60
TTA	TAACCGT CTTTCCCTT ATG CTA AGG ATA GO Met Leu Arg Ile Al -15	CC CTT ACA CTC ATC CC la Leu Thr Leu Ile Pr -10	
ATO Met -5	CTG TCA AGG GCT GCT GGT TGG TGC TGG Leu Ser Arg Ala Ala Gly Trp Cys Trp 1	G TAC AAG GAG CCC ACT O Tyr Lys Glu Pro Thr 10	Gln
	TTT TCT TAC CTT TGC CTG CCG GGG Phe Ser Tyr Leu Cys Leu Pro Gly 15 20		187
(2)	INFORMATION FOR SEQ ID NO: 180:		
	(i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 145 base pairs (B) TYPE: NUCLEIC ACID (C) STRANDEDNESS: DOUBLE (D) TOPOLOGY: LINEAR		
	(ii) MOLECULE TYPE: CDNA		•
	<pre>(vi) ORIGINAL SOURCE: (A) ORGANISM: Homo Sapiens (F) TISSUE TYPE: Lung</pre>		
	regio	lastn Lity 97 on 1654 Al29105	
	<pre>(ix) FEATURE: (A) NAME/KEY: sig_peptide (B) LOCATION: 2379 (C) IDENTIFICATION METHOD: Vo (D) OTHER INFORMATION: score seq I</pre>		
	(xi) SEQUENCE DESCRIPTION: SEQ ID	NO: 180:	
ACG	GCAGAAG GGACGTCCCC AA ATG ACA CTC GO Met Thr Leu Gl	GA GGC CGC CTC CCT GG Ly Gly Arg Leu Pro Gl -15	
CG9 Arg	TGC TCG GTG CCG GGA GTA GCG GCA CGC Cys Ser Val Pro Gly Val Ala Ala Arc -5	G CTT TCT ACC CCG CCT g Leu Ser Thr Pro Pro l 5	CAG 100
GTG Val	CGC CAG CAC GTT TTC TGG GCA GCA TCC Arg Gln His Val Phe Trp Ala Ala Ser 10	r GTG TGT DAG GMA ACG r Val Cys Xaa Xaa Thr 20	145

(2) INFORMATION FOR SEQ ID NO: 181:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 289 base pairs
 - (B) TYPE: NUCLEIC ACID
 - (C) STRANDEDNESS: DOUBLE
 - (D) TOPOLOGY: LINEAR
- (ii) MOLECULE TYPE: CDNA
- (vi) ORIGINAL SOURCE:
 - (A) ORGANISM: Homo Sapiens
 - (F) TISSUE TYPE: Liver
- (ix) FEATURE:
 - (A) NAME/KEY: other
 - (B) LOCATION: 23..291
 - (C) IDENTIFICATION METHOD: blastn
 - (D) OTHER INFORMATION: identity 97

region 1..269 id T29966

est

- (ix) FEATURE:
 - (A) NAME/KEY: other
 - (B) LOCATION: 123..291
 - (C) IDENTIFICATION METHOD: blastn
 - (D) OTHER INFORMATION: identity 97

region 100..268

· id HSC1VD021

est

- (ix) FEATURE:
 - (A) NAME/KEY: other
 - (B) LOCATION: 23..124
 - (C) IDENTIFICATION METHOD: blastn
 - (D) OTHER INFORMATION: identity 96

region 1..102

id HSC1VD021

est

- (ix) FEATURE:
 - (A) NAME/KEY: other
 - (B) LOCATION: 56..291
 - (C) IDENTIFICATION METHOD: blastn
 - (D) OTHER INFORMATION: identity 97

region 1..236

id T35162

est

- (ix) FEATURE:
 - (A) NAME/KEY: other
 - (B) LOCATION: 122..291
 - (C) IDENTIFICATION METHOD: blastn
 - (D) OTHER INFORMATION: identity 97

region 41..210

id R12259

est

164

PCT/IB98/01233

								10	4						
	(ix)	FEATU (A) N (B) L (C) I (D) O	AME/F OCATI DENTI	ON: FICE	123. ATIO	25	THOD:	dent	ity 9 n 75		3				
	(ix)	FEATU (A) N (B) L (C) I (D) O	AME/F OCATI DENTI	ON:	48. ATION	124 N ME:	: io re io	dent	ity 9 n 1.						
	(ix)	FEATU (A) N. (B) L (C) I (D) O	AME/F OCATI DENTI	ON:	.68 1017	130 1 ME:	HOD:	core							
	(xi)	SEQUE	NCE [ESCF	RIPT	ON:	SEQ	ID	NO: 1	181:					
AGT	AGGAAS	G CGCC	GSCCG	T GO	GAGG	CGCC	A CGT	rccc:	TGC	SGC	GCG	GGA (SAGAI	MATCGC	60
rtg(SACTIC	G GGGC	GGCCT	C GG	GACGO		ATG (Chr 1					112
CTG Leu	CAG GG Gln A	CA GCC la Ala	CTG Leu	CTC Leu	TGC Cys 1	GTC Val	AAC Asn	GCC Ala	ATC Ile 5	GCA Ala	GTG Val	CTG Leu	CAC His	GAG Glu 10	160
GAG Glu	CGA T' Arg Pi	TC CTC he Leu	AAG Lys 15	AAC Asn	ATT Ile	GGC Gly	TGG Trp	GGA Gly 20	ACA Thr	GAC Asp	CAG Gln	GGA Gly	ATT Ile 25	GGT Gly	208
GGA Gly	TTT G	GA GAA ly Glu 30	GAG Glu	CCG Pro	GGA Gly	ATT Ile	AAA Lys 35	TCA Ser	SAG Xaa	STA Xaa	ATG Met	AVS Xaa 40	CTT Leu	ATT Ile	256
	Ser V	TA AGA al Arg 45													289
(2)	INFOR	MATION	FOR	SEQ	ID t	10:	182:								
	(i)	SEQUEN (A) L (B) T (C) S (D) T	ENGTI YPE : TRANI	i: 39 NUCI DEDNE	56 ba LEIC ESS:	ase p ACII DOU!	pair: D	S	,						
	(ii)	MOLEC	ULE 1	YPE:	: CDI	NA									

```
(vi) ORIGINAL SOURCE:
         (A) ORGANISM: Homo Sapiens
          (F) TISSUE TYPE: Pancreas
     (ix) FEATURE:
          (A) NAME/KEY: other
          (B) LOCATION: 243..358
          (C) IDENTIFICATION METHOD: blastn
          (D) OTHER INFORMATION: identity 98
                                 region 99..214
                                  id R74138
                                  est
    (ix) FEATURE:
         (A) NAME/KEY: other
          (B) LOCATION: 2..99
          (C) IDENTIFICATION METHOD: blastn
          (D) OTHER INFORMATION: identity 92
                                 region 2..99
                                 id R74138
     (ix) FEATURE:
          (A) NAME/KEY: other
          (B) LOCATION: 243..358
          (C) IDENTIFICATION METHOD: blastn
          (D) OTHER INFORMATION: identity 98
                                  region 68..183
                                  id C18563
                                  est .
     (ix) FEATURE:
          (A) NAME/KEY: other
          (B) LOCATION: 32..99
          (C) IDENTIFICATION METHOD: blastn
          (D) OTHER INFORMATION: identity 94
                                  region 1..68
                                 id C18563
                                  est
     (ix) FEATURE:
          (A) NAME/KEY: sig_peptide
          (B) LOCATION: 81..143
          (C) IDENTIFICATION METHOD: Von Heijne matrix
          (D) OTHER INFORMATION: score 3.5
                                  seq LKVVFMVFASLXA/WY
     (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 182:
ACTTCCTGAC CCAGGGGCTC CGCTGGCTGC GGTCGCCTGG GAKYTGCCGC CAGGGCCAGG
AGGGGAKYGG CACCTGGAAG ATG CGC CCA TTG GCT GGT GGC CTG CTC AAG GTG 113
                      Met Arg Pro Leu Ala Gly Gly Leu Leu Lys Val
GTG TTC ATG GTC TTC GCC TCC TTG KRW GCC TGG TAT TCG GGG TAC CTG
                                                                    161
Val Phe Met Val Phe Ala Ser Leu Xaa Ala Trp Tyr Ser Gly Tyr Leu
```

. 166	
CTC GCA GAS NTC ATT CCA GAT GCA CCC CTG TCC AGT GCT GCC TAT AG Leu Ala Xaa Xaa Ile Pro Asp Ala Pro Leu Ser Ser Ala Ala Tyr Se 10 15 20	C 209
ATC CGC AGC ATC GGG GAG AGG CCT GTC CTC AAA GCT CCA GTC CCC AA Ile Arg Ser Ile Gly Glu Arg Pro Val Leu Lys Ala Pro Val Pro Ly 25 30 35	
AGG CAA AAA TGT GAC CAC TGG ACT CCC TGC CCA TCT GAS RCC TAT GC Arg Gln Lys Cys Asp His Trp Thr Pro Cys Pro Ser Xaa Xaa Tyr Al 40 45 50	
TAC AGG TTA CTC AGC GGA GGT GGC AGA AGC AAG TAC GCC AAA ATC TG Tyr Arg Leu Leu Ser Gly Gly Gly Arg Ser Lys Tyr Ala Lys Ile Cy 55 60 65 7	s
TTT Phe	356
(2) INFORMATION FOR SEQ ID NO: 183: (i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 289 base pairs (B) TYPE: NUCLEIC ACID (C) STRANDEDNESS: DOUBLE (D) TOPOLOGY: LINEAR (ii) MOLECULE TYPE: CDNA (vi) ORIGINAL SOURCE: (A) ORGANISM: Homo Sapiens (F) TISSUE TYPE: Colon (ix) FEATURE: (A) NAME/KEY: other (B) LOCATION: complement (207287) (C) IDENTIFICATION METHOD: blastn (D) OTHER INFORMATION: identity 91 region 220300 id H49636	
est (ix) FEATURE: (A) NAME/KEY: sig_peptide (B) LOCATION: 116181 (C) IDENTIFICATION METHOD: Von Heijne matrix (D) OTHER INFORMATION: score 3.5 seq AAPVAAGLGPVIS/RP	
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 183:	
AGTCACGCTT TGGTTTCCGG GTCGGTTCTG GCAGGTCTGA GCGCTCCGAC TTCCAGA	GGA 60
GCGCTGTGCA CGTGGAGAAG AGCGGGGACT CGGCGACCCT GCCCTCCGA CCCTC A	TG 118
TTC GAA GAG CCT GAG TGG GCC GAG GCC CCA GTA GCC GCG GGC CT Phe Glu Glu Pro Glu Trp Ala Glu Ala Ala Pro Val Ala Ala Glu Le	

167

(2) INFORMATION FOR SEQ ID NO: 184:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 139 base pairs
 - (B) TYPE: NUCLEIC ACID
 - (C) STRANDEDNESS: DOUBLE
 - (D) TOPOLOGY: LINEAR
- (ii) MOLECULE TYPE: CDNA
- (vi) ORIGINAL SOURCE:
 - (A) ORGANISM: Homo Sapiens
 - (F) TISSUE TYPE: Lung (cells)
- (ix) FEATURE:
 - (A) NAME/KEY: other
 - (B) LOCATION: 34..129
 - (C) IDENTIFICATION METHOD: blastn
 - (D) OTHER INFORMATION: identity 96 region 1..96 id W54807
- (ix) FEATURE:
 - (A) NAME/KEY: sig_peptide
 - (B) LOCATION: 50..118
 - (C) IDENTIFICATION METHOD: Von Heijne matrix
 - (D) OTHER INFORMATION: score 3.5 seq ALYNIIYVCGIOG/IT
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 184:
- ACCGTTGTCA CCAGCCTTGT TTCTAGTGTG TATATATGTC GCCCCCGTG ATG CAT ATA 58

 Met His Ile
- TAC ACA GGT ATT AAA TAT ATC GCT CTA TAT AAT ATT ATA TAT GTG TGT

 Tyr Thr Gly Ile Lys Tyr Ile Ala Leu Tyr Asn Ile Ile Tyr Val Cys

 -20 -15 -5
- GGT ATC CAA GGA ATC ACT TTT ATG AGG GCA CGG Gly Ile Gln Gly Ile Thr Phe Met Arg Ala Arg

PCT/IB98/01233

- (2) INFORMATION FOR SEQ ID NO: 185:
 - (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 96 amino acids
 - (B) TYPE: AMINO ACID
 - (D) TOPOLOGY: LINEAR
 - (ii) MOLECULE TYPE: PROTEIN
 - (vi) ORIGINAL SOURCE:
 - (A) ORGANISM: Homo Sapiens
 - (D) DEVELOPMENTAL STAGE: Fetal
 - (F) TISSUE TYPE: liver
 - (ix) FEATURE:
 - (A) NAME/KEY: sig_peptide
 - (B) LOCATION: -20..-1
 - (C) IDENTIFICATION METHOD: Von Heijne matrix
 - (D) OTHER INFORMATION: score 15.8

seq LLLLLLRHGAQG/KP

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 185:

Met Met Trp Arg Pro Ser Val Leu Leu Leu Leu Leu Leu Leu Leu Arg His
-20 -15 -10 -5

Gly Ala Gln Gly Lys Pro Ser Pro Asp Ala Gly Pro His Gly Gln Gly $1 \hspace{1cm} 5 \hspace{1cm} 10$

Arg Val His Gln Ala Ala Pro Leu Ser Asp Ala Pro His Asp Asp Ala 15 20 25

His Gly Asn Phe Gln Tyr Asp His Glu Ala Phe Leu Gly Arg Glu Val 30 35 40

Ala Lys Glu Phe Asp Gln Leu Thr Pro Glu Glu Ser Gln Ala Arg Leu
45 50 55 60

Gly Arg Ile Val Asp Arg Met Asp Arg Xaa Gly Thr Ala Thr Ala Gly
65 70 75

- (2) INFORMATION FOR SEQ ID NO: 186:
 - (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 30 amino acids
 - (B) TYPE: AMINO ACID
 - (D) TOPOLOGY: LINEAR
 - (ii) MOLECULE TYPE: PROTEIN
 - (vi) ORIGINAL SOURCE:
 - (A) ORGANISM: Homo Sapiens
 - (F) TISSUE TYPE: Lung (cells)
 - (ix) FEATURE:
 - (A) NAME/KEY: sig_peptide
 - (B) LOCATION: -28..-1

169

- (C) IDENTIFICATION METHOD: Von Heijne matrix
 - (D) OTHER INFORMATION: score 11.4 seq LLLILFLYGLCSG/WR
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 186:

Met Gly Lys Ile Cys Lys Asn Trp Val Ser Phe Leu Asp Asn Val Leu
-25 -20 -15

Leu Leu Ile Leu Phe Leu Tyr Gly Leu Cys Ser Gly Trp Arg

- (2) INFORMATION FOR SEQ ID NO: 187:
 - (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 52 amino acids
 - (B) TYPE: AMINO ACID
 - (D) TOPOLOGY: LINEAR
 - (ii) MOLECULE TYPE: PROTEIN
 - (vi) ORIGINAL SOURCE:
 - (A) ORGANISM: Homo Sapiens
 - (F) TISSUE TYPE: Large intestine
 - (ix) FEATURE:
 - (A) NAME/KEY: sig_peptide
 - (B) LOCATION: -18..-1
 - (C) IDENTIFICATION METHOD: Von Heijne matrix
 - (D) OTHER INFORMATION: score 11.1 seq LLALLCASASGNA/IQ
 - (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 187:

Met Leu Thr Val Ala Leu Leu Ala Leu Leu Cys Ala Ser Ala Ser Gly
-15 -10 -5

Asn Ala Ile Gln Ala Arg Ser Ser Ser Tyr Ser Gly Glu Tyr Gly Leu

1 5 . 10

Val Val Glu Ser Asp Ser Leu Ile Leu Ala Thr Ser Trp Thr Ala Pro
15 20 25 30

Ser Pro Pro Thr

- (2) INFORMATION FOR SEQ ID NO: 188:
 - (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 32 amino acids
 - (B) TYPE: AMINO ACID
 - (D) TOPOLOGY: LINEAR
 - (ii) MOLECULE TYPE: PROTEIN
 - (vi) ORIGINAL SOURCE:

170

- (A) ORGANISM: Homo Sapiens
- (F) TISSUE TYPE: Pancreas
- (ix) FEATURE:
 - (A) NAME/KEY: sig_peptide
 - (B) LOCATION: -13..-1
 - (C) IDENTIFICATION METHOD: Von Heijne matrix
 - (D) OTHER INFORMATION: score 10.4

seq MVLLLCLSCLIFS/CL

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 188:

Met Val Leu Leu Cys Leu Ser Cys Leu Ile Phe Ser Cys Leu Thr
-10 -5 1

Phe Ser Trp Leu Lys Ile Trp Gly Lys Met Thr Asp Ser Lys Pro Met
5 10

- (2) INFORMATION FOR SEQ ID NO: 189:
 - (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 90 amino acids
 - (B) TYPE: AMINO ACID
 - (D) TOPOLOGY: LINEAR
 - (ii) MOLECULE TYPE: PROTEIN
 - (vi) ORIGINAL SOURCE:
 - (A) ORGANISM: Homo Sapiens
 - (F) TISSUE TYPE: Liver
 - (ix) FEATURE:
 - (A) NAME/KEY: sig_peptide
 - (3) LOCATION: -21..-1
 - (C) IDENTIFICATION METHOD: Von Heijne matrix
 - (D) OTHER INFORMATION: score 10

seq LWALAMVTRPASA/AP

- (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 189:
- Met Pro Val Pro Ala Leu Cys Leu Leu Trp Ala Leu Ala Met Val Thr
 -20 -15 -10
- Arg Pro Ala Ser Ala Ala Pro Met Gly Gly Pro Glu Leu Ala Gln His -5 10
- Glu Glu Leu Thr Leu Leu Phe His Gly Thr Leu Gln Leu Gly Gln Ala 15 20 25
- Leu Asn Gly Val Tyr Arg Thr Thr Glu Gly Arg Leu Thr Lys Ala Arg 30 35 40
- Asn Ser Leu Gly Leu Tyr Gly Arg Thr Ile Glu Leu Leu Gly Gln Glu
 45 55
- Val Ser Arg Gly Arg Asp Ala Ala Gln Gly

- (2) INFORMATION FOR SEQ ID NO: 190:
 - (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 44 amino acids
 - (B) TYPE: AMINO ACID
 - (D) TOPOLOGY: LINEAR
 - (ii) MOLECULE TYPE: PROTEIN
 - (vi) ORIGINAL SOURCE:
 - (A) ORGANISM: Homo Sapiens
 - (F) TISSUE TYPE: Lung (cells)
 - (ix) FEATURE:
 - (A) NAME/KEY: sig_peptide
 - (B) LOCATION: -41..-1
 - (C) IDENTIFICATION METHOD: Von Heijne matrix
 - (D) OTHER INFORMATION: score 9.4

seq LLPLWVFLPLSLG/PP

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 190:

Met His Leu Arg Gly Ser His Thr Tyr Pro Ser Cys Pro Ser Ser Glu
-40 -35 -30

Leu Arg Leu Asp Ser Leu Trp Gln His His Arg Gln Leu Leu Pro Leu -25 -15 -10

Trp Val Phe Leu Pro Leu Ser Leu Gly Pro Pro Gly
-5

- (2) INFORMATION FOR SEQ ID NO: 191:
 - (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 77 amino acids
 - (B) TYPE: AMINO ACID
 - (D) TOPOLOGY: LINEAR
 - (ii) MOLECULE TYPE: PROTEIN
 - (vi) ORIGINAL SOURCE:
 - (A) ORGANISM: Homo Sapiens
 - (F) TISSUE TYPE: Lung
 - (ix) FEATURE:
 - (A) NAME/KEY: sig_peptide
 - (B) LOCATION: -26..-1
 - (C) IDENTIFICATION METHOD: Von Heijne matrix
 - (D) OTHER INFORMATION: score 9.2

seq FLLMTLLLGGLTG/VA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 191:

Met Pro Val Pro Ala Ser Trp Pro His Leu Pro Ser Pro Phe Leu Leu

172

-25 **-**20 -15

Met Thr Leu Leu Gly Gly Leu Thr Gly Val Ala Xaa Glu Glu Glu -10 -5 1 5

Leu Gln Val Xaa Gln Pro Asp Lys Ser Ile Ser Val Ala Ala Gly Lys
10 15 20

Xaa Ala Thr Leu His Cys Thr Val Thr Xaa Leu Ile Xaa Val Gly Pro 25 30 35

Ile Gln Trp Xaa Arg Gly Ala Gly Pro Gly Arg Glu Leu
40 45 50

(2) INFORMATION FOR SEQ ID NO: 192:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 93 amino acids
 - (B) TYPE: AMINO ACID
 - (D) TOPOLOGY: LINEAR
- (ii) MOLECULE TYPE: PROTEIN
- (vi) ORIGINAL SOURCE:
 - (A) ORGANISM: Homo Sapiens
 - (F) TISSUE TYPE: Large intestine
- (ix) FEATURE:
 - (A) NAME/KEY: sig_peptide
 - (B) LOCATION: -23..-1
 - (C) IDENTIFICATION METHOD: Von Heijne matrix
 - (D) OTHER INFORMATION: score 9.1 seq LVAMLLLVFPTVS/RS
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 192:

Met Ala Gln Arg Cys Val Cys Val Leu Ala Leu Val Ala Met Leu Leu -20 -15 -10

Leu Val Phe Pro Thr Val Ser Arg Ser Met Gly Pro Arg Ser Gly Glu -5 5

His Gln Arg Ala Ser Arg Ile Pro Ser Gln Phe Ser Lys Glu Glu Arg 10 15 20 25

Val Ala Met Lys Glu Ala Leu Lys Gly Ala Ile Gln Ile Pro Thr Val 30 35 40

Thr Phe Ser Ser Glu Lys Ser Asn Thr Thr Ala Leu Ala Glu Phe Gly
45 50 55

Asn Thr Phe Ile Lys Ser Phe Leu Gln Trp Ser Ala Pro 60 65 70

(2) INFORMATION FOR SEQ ID NO: 193:

173

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 19 amino acids
- (B) TYPE: AMINO ACID
- (D) TOPOLOGY: LINEAR
- (ii) MOLECULE TYPE: PROTEIN
- (vi) ORIGINAL SOURCE:
 - (A) ORGANISM: Homo Sapiens
 - (F) TISSUE TYPE: Lung (cells)
- (ix) FEATURE:
 - (A) NAME/KEY: sig peptide
 - (B) LOCATION: -16..-1
 - (C) IDENTIFICATION METHOD: Von Heijne matrix
 - (D) OTHER INFORMATION: score 9

seq LISFLLLLLLLP/AR

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 193:

Met Asp Tyr Leu Ile Ser Phe Leu Leu Leu Leu Leu Leu Leu Leu Leu Pro
-15 -10 -5

Ala Arg Gly

- (2) INFORMATION FOR SEQ ID NO: 194:
 - (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 88 amino acids
 - (B) TYPE: AMINO ACID
 - (D) TOPOLOGY: LINEAR
 - (ii) MOLECULE TYPE: PROTEIN
 - (vi) ORIGINAL SOURCE:
 - (A) ORGANISM: Homo Sapiens
 - (F) TISSUE TYPE: Colon
 - (ix) FEATURE:
 - (A) NAME/KEY: sig_peptide
 - (B) LOCATION: -84..-1
 - (C) IDENTIFICATION METHOD: Von Heijne matrix
 - (D) OTHER INFORMATION: score 8.9

seq LLGVLGIFGLTFA/FI

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 194:

Met Ala Thr Thr Val Pro Asp Gly Cys Arg Asn Gly Leu Lys Ser Lys
-80 -75 -70

Tyr Tyr Arg Leu Cys Asp Lys Ala Glu Ala Trp Gly Ile Val Leu Glu
-65 -60 -55

Thr Val Ala Thr Ala Gly Val Val Thr Ser Val Ala Phe Met Leu Thr
-50 -45 -40

174

Leu Pro Ile Leu Val Cys Lys Val Gln Asp Ser Asn Arg Arg Lys Met
-35 -25

Leu Pro Thr Gln Phe Leu Phe Leu Gly Val Leu Gly Ile Phe Gly -20 -15 -10 -5

Leu Thr Phe Ala Phe Ile Ile Gly

- (2) INFORMATION FOR SEQ ID NO: 195:
 - (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 71 amino acids
 - (B) TYPE: AMINO ACID
 - (D) TOPOLOGY: LINEAR
 - (ii) MOLECULE TYPE: PROTEIN
 - (vi) ORIGINAL SOURCE:
 - (A) ORGANISM: Homo Sapiens
 - (F) TISSUE TYPE: Large intestine
 - (ix) FEATURE:
 - (A) NAME/KEY: sig_peptide
 - (B) LOCATION: -19..-1
 - (C) IDENTIFICATION METHOD: Von Heijne matrix
 - (D) OTHER INFORMATION: score 8.9 seq WLFLVIFIKGVQC/QE
 - (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 195:

Met Glu Ser Gly Leu Ser Trp Leu Phe Leu Val Ile Phe Ile Lys Gly
-15 -10 -5

Val Gln Cys Gln Glu Gln Leu Val Glu Ser Gly Gly Gly Val Val Lys

1 5 10

Pro Gly Gly Ser Leu Arg Leu Ser Cys Glu Ala Ser Gly Phe Thr Phe 15 20 25

Ser Asp Phe Xaa Met Met Trp Ile Arg Gln Thr Pro Gly Lys Gly Leu 30 40 45

Glu Tyr Val Gly Ile His Gln 50

- (2) INFORMATION FOR SEQ ID NO: 196:
 - (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 21 amino acids
 - (B) TYPE: AMINO ACID
 - (D) TOPOLOGY: LINEAR
 - (ii) MOLECULE TYPE: PROTEIN
 - (vi) ORIGINAL SOURCE:

175

- (A) ORGANISM: Homo Sapiens
- (F) TISSUE TYPE: Lung (cells)
- (ix) FEATURE:
 - (A) NAME/KEY: sig_peptide
 - (B) LOCATION: -18..-1
 - (C) IDENTIFICATION METHOD: Von Heijne matrix
 - (D) OTHER INFORMATION: score 8.6

seq VLLHVAFLPGRFG/RP

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 196:

Met Ser Gly Thr Ser Val Leu Leu His Val Ala Phe Leu Pro Gly Arg
-15 -10 -5

Phe Gly Arg Pro Leu

- (2) INFORMATION FOR SEQ ID NO: 197:
 - (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 53 amino acids
 - (B) TYPE: AMINO ACID
 - (D) TOPOLOGY: LINEAR
 - (ii) MOLECULE TYPE: PROTEIN
 - (vi) ORIGINAL SOURCE:
 - (A) ORGANISM: Homo Sapiens
 - (F) TISSUE TYPE: Lung (cells)
 - (ix) FEATURE:
 - (A) NAME/KEY: sig_peptide
 - (B) LOCATION: -17..-1
 - (C) IDENTIFICATION METHOD: Von Heijne matrix
 - (D) OTHER INFORMATION: score 8.6

seq LLPVSLLLSVAVS/AI

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 197:

Met Leu Gln Gly Leu Leu Pro Val Ser Leu Leu Leu Ser Val Ala Val
-15 -5

Ser Ala Ile Lys Glu Leu Pro Gly Val Lys Lys Tyr Glu Val Val Tyr
1 5 10 15

Pro Ile Arg Leu His Pro Leu His Lys Arg Glu Ala Lys Glu Pro Glu 20 25 30

Gln Gln Glu Arg Arg 35

- (2) INFORMATION FOR SEQ ID NO: 198:
 - (i) SEQUENCE CHARACTERISTICS:

176

(A) LENGTH: 30 amino acids

(B) TYPE: AMINO ACID

(D) TOPOLOGY: LINEAR

- (ii) MOLECULE TYPE: PROTEIN
- (vi) ORIGINAL SOURCE:
 - (A) ORGANISM: Homo Sapiens
 - (F) TISSUE TYPE: Colon
- (ix) FEATURE:
 - (A) NAME/KEY: sig_peptide
 - (B) LOCATION: -15..-1
 - (C) IDENTIFICATION METHOD: Von Heijne matrix
 - (D) OTHER INFORMATION: score 8.6

seq ICHVSLLLQLCSS/CK

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 198:

Met His Ile Cys His Val Ser Leu Leu Leu Gln Leu Cys Ser Ser Cys
-15 -10 -5 1

Lys Lys Ser Pro Leu Lys Leu Leu Gln Lys Ala Gln Arg 5 10 15

- (2) INFORMATION FOR SEQ ID NO: 199:
 - (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 37 amino acids
 - (B) TYPE: AMINO ACID
 - (D) TOPOLOGY: LINEAR
 - (ii) MOLECULE TYPE: PROTEIN
 - (vi) ORIGINAL SOURCE:
 - (A) ORGANISM: Homo Sapiens
 - (F) TISSUE TYPE: Liver
 - (ix) FEATURE:
 - (A) NAME/KEY: sig_peptide
 - (B) LOCATION: -33..-1
 - (C) IDENTIFICATION METHOD: Von Heijne matrix
 - (D) OTHER INFORMATION: score 8.2

seg LLCSLLFSFPFLC/HP

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 199:

Met Ile Phe Ala Asp Arg Thr His Ser Ser Ala Phe Thr Leu Met Arg
-30 -25 -20

Ser Tyr Ser Leu Leu Cys Ser Leu Leu Phe Ser Phe Pro Phe Leu
-15 -10 -5

Cys His Pro Leu Arg

177

(2) INFORMATION FOR SEQ ID NO: 200:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 72 amino acids
 - (B) TYPE: AMINO ACID
 - (D) TOPOLOGY: LINEAR
- (ii) MOLECULE TYPE: PROTEIN
- (vi) ORIGINAL SOURCE:
 - (A) ORGANISM: Homo Sapiens
 - (F) TISSUE TYPE: Pancreas
- (ix) FEATURE:
 - (A) NAME/KEY: sig_peptide
 - (B) LOCATION: -20..-1
 - (C) IDENTIFICATION METHOD: Von Heijne matrix
 - (D) OTHER INFORMATION: score 8

seq VCVLVGSFSASLA/GT

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 200:

Met Ala Phe Leu Pro Ser Trp Val Cys Val Leu Val Gly Ser Phe Ser -20 -15 -10 -5

Ala Ser Leu Ala Gly Thr Ser Asn Leu Ser Glu Thr Glu Pro Pro Leu
1 5 10

Trp Lys Glu Ser Pro Gly Gln Leu Ser Asp Tyr Arg Val Glu Asn Ser 15 20 25

Met Tyr Ile Ile Asn Pro Trp Val Tyr Leu Glu Arg Met Gly Met Tyr 30 35 40

Lys Ile Ile Leu Asn Gln Thr Lys 45 50

- (2) INFORMATION FOR SEQ ID NO: 201:
 - (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 26 amino acids
 - (B) TYPE: AMINO ACID
 - (D) TOPOLOGY: LINEAR
 - (ii) MOLECULE TYPE: PROTEIN
 - (vi) ORIGINAL SOURCE:
 - (A) ORGANISM: Homo Sapiens
 - (F) TISSUE TYPE: Lung (cells)
 - (ix) FEATURE:
 - (A) NAME/KEY: sig_peptide
 - (B) LOCATION: -14..-1
 - (C) IDENTIFICATION METHOD: Von Heijne matrix
 - (D) OTHER INFORMATION: score 7.7

seq FLVSCVICTGSFA/FN

178

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 201:

Met Phe Leu Val Ser Cys Val Ile Cys Thr Gly Ser Phe Ala Phe Asn -10 -5 1

Asn Ser Asn Val Pro Leu Pro Ser Ser Arg

- (2) INFORMATION FOR SEQ ID NO: 202:
 - (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 68 amino acids
 - (B) TYPE: AMINO ACID
 - (D) TOPOLOGY: LINEAR
 - (ii) MOLECULE TYPE: PROTEIN
 - (vi) ORIGINAL SOURCE:
 - (A) ORGANISM: Homo Sapiens
 - (F) TISSUE TYPE: Lung (cells)
 - (ix) FEATURE:
 - (A) NAME/KEY: sig_peptide
 - (B) LOCATION: -44..-1
 - (C) IDENTIFICATION METHOD: Von Heijne matrix
 - (D) OTHER INFORMATION: score 7.5

seq LMIPLLLTPITA/TS

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 202:

Met Lys Lys Thr Gly Asp Gly Gly Thr Leu Ser Thr Glu Arg Ile Gly
-40 -35 -30

Gly Ala Ala Leu Leu Ser Leu Leu Leu Lys Arg Met Lys Met Thr Leu
-25 -20 -15

Met Ile Pro Leu Leu Leu Thr Pro Ile Thr Ala Thr Ser Xaa Ser
-10 -5 1

Arg Trp Pro Glu Ile Gly Val Val Ala Ile Arg Ser Gln Leu Arg Ala 5 10 15 20

Leu His Thr Cys

- (2) INFORMATION FOR SEQ ID NO: 203:
 - (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 116 amino acids
 - (B) TYPE: AMINO ACID
 - (D) TOPOLOGY: LINEAR
 - (ii) MOLECULE TYPE: PROTEIN
 - (vi) ORIGINAL SOURCE:
 - (A) ORGANISM: Homo Sapiens

- (F) TISSUE TYPE: Lung (cells)
- (ix) FEATURE:
 - (A) NAME/KEY: sig_peptide
 - (B) LOCATION: -25..-1
 - (C) IDENTIFICATION METHOD: Von Heijne matrix
 - (D) OTHER INFORMATION: score 7.5

seq XILLAGWCPDTRA/GG

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 203:

Met Gly Phe Phe Leu Pro His Gly Ile Ser Asp Ala Xaa Ile Leu Leu -25 -15 -10

Ala Gly Trp Cys Pro Asp Thr Arg Ala Gly Gly Trp Ala Asp Leu Cys
-5 1 5

Leu Pro Glu Asn Arg Gly Pro Lys Pro Pro Ser Pro Arg Ser Ala Leu 10 15 20

Gly Ser Gly Arg Gly Leu Gly Ser Gly Gln Pro Glu Val Glu Pro Pro 25 30 35

Ala Pro Glu Gln Ala Trp Glu Ser Leu Gln Gly Gly Leu Gly Xaa Cys 40 45 50 55

Ser Xaa Ala Arg Pro Ser Pro Gly Phe Trp Ala Arg Ala Ser Leu Ala 60 65 70

Val Gly Ala Gly Xaa Val Gly Gly Thr Leu Leu Asn Trp Glu Ile Ala 75 80 85

Ser Asp Leu Gln 90

- (2) INFORMATION FOR SEQ ID NO: 204:
 - (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 73 amino acids
 - (B) TYPE: AMINO ACID
 - (D) TOPOLOGY: LINEAR
 - (ii) MOLECULE TYPE: PROTEIN
 - (vi) ORIGINAL SOURCE:
 - (A) ORGANISM: Homo Sapiens
 - (F) TISSUE TYPE: Colon
 - (ix) FEATURE:
 - (A) NAME/KEY: sig_peptide
 - (B) LOCATION: -46..-1
 - (C) IDENTIFICATION METHOD: Von Heijne matrix
 - (D) OTHER INFORMATION: score 7.5 seq VCTLLSSHPASRC/RP
 - (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 204:

180

Met Trp Leu Arg Pro Gly Ser Cys Trp Ser Thr Arg Glu Pro Arg Arg
-45 -40 -35

Ala Pro Arg Thr Ser Ala Ser Ser Leu Ser Ser Phe Leu Gly Pro Ser -30 -25 -20 -15

Ala Val Cys Thr Leu Leu Ser Ser His Pro Ala Ser Arg Cys Arg Pro
-10 -5 1

Ser Thr Phe Leu Ala Pro Gly Phe Cys Ile Cys Pro Ser His Cys Leu 5 10 15

Ser Cys Ala Asn Ala Thr Asp Pro Ala 20 25

- (2) INFORMATION FOR SEQ ID NO: 205:
 - (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 48 amino acids
 - (B) TYPE: AMINO ACID
 - (D) TOPOLOGY: LINEAR
 - (ii) MOLECULE TYPE: PROTEIN
 - (vi) ORIGINAL SOURCE:
 - (A) ORGANISM: Homo Sapiens
 - (F) TISSUE TYPE: Lung
 - (ix) FEATURE:
 - (A) NAME/KEY: sig_peptide
 - (B) LOCATION: -39..-1
 - (C) IDENTIFICATION METHOD: Von Heijne matrix
 - (D) OTHER INFORMATION: score 7.4 seq LFLLSLFCRLYHG/TI
 - (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 205:

Met Ser Glu Gly Met Val Thr Leu Leu Thr Phe Ser Cys Leu Trp Thr -35 -30 -25

Asp Asp Ser Phe Met Ser Xaa Leu Asn Val Leu Phe Leu Leu Ser Leu -20 -15 -10

Phe Cys Arg Leu Tyr His Gly Thr Ile Phe Phe Leu Leu Ala Leu Leu -5 5

- (2) INFORMATION FOR SEQ ID NO: 206:
 - (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 53 amino acids
 - (B) TYPE: AMINO ACID
 - (D) TOPOLOGY: LINEAR
 - (ii) MOLECULE TYPE: PROTEIN
 - (vi) ORIGINAL SOURCE:

181

- (A) ORGANISM: Homo Sapiens
- (F) TISSUE TYPE: Lung
- (ix) FEATURE:
 - (A) NAME/KEY: sig_peptide
 - (B) LOCATION: -14..-1
 - (C) IDENTIFICATION METHOD: Von Heijne matrix
 - (D) OTHER INFORMATION: score 7.3

seg LILGLPLCRPLWI/QR

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 206:

Met Leu Ile Leu Gly Leu Pro Leu Cys Arg Pro Leu Trp Ile Gln Arg

Ala Ala Ala Pro Phe Val Leu Trp Ala Trp Leu Trp Ala Arg Ser 10 15

Ser Thr Ser Leu Gly Arg Pro Pro Phe Leu Pro Arg Leu Leu Pro Ser 20 25 30

Pro Pro Asp Pro Glu

- (2) INFORMATION FOR SEQ ID NO: 207:
 - (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 49 amino acids
 - (B) TYPE: AMINO ACID
 - (D) TOPOLOGY: LINEAR
 - (ii) MOLECULE TYPE: PROTEIN
 - (vi) ORIGINAL SOURCE:
 - (A) ORGANISM: Homo Sapiens
 - (F) TISSUE TYPE: Lung
 - (ix) FEATURE:
 - (A) NAME/KEY: sig_peptide
 - (B) LOCATION: -25..-1
 - (C) IDENTIFICATION METHOD: Von Heijne matrix
 - (D) OTHER INFORMATION: score 7.3

seq HFILLVLPCLIFS/HF

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 207:

Met Tyr Ile Tyr Phe Phe Val Leu Cys Xaa Leu Ser His Phe Ile Leu -25 -15 -10

Leu Val Leu Pro Cys Leu Ile Phe Ser His Phe Thr Leu Phe Leu Phe -5 1 5

Tyr Ser Ala Leu Leu Asp Ile Pro Leu Phe Phe Lys Tyr Ser Leu Ile 10 15 20

Glu

- (2) INFORMATION FOR SEQ ID NO: 208:
 - (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 69 amino acids
 - (B) TYPE: AMINO ACID
 - (D) TOPOLOGY: LINEAR
 - (ii) MOLECULE TYPE: PROTEIN
 - (vi) ORIGINAL SOURCE:
 - (A) ORGANISM: Homo Sapiens
 - (F) TISSUE TYPE: Lung
 - (ix) FEATURE:
 - (A) NAME/KEY: sig_peptide
 - (B) LOCATION: -42..-1
 - (C) IDENTIFICATION METHOD: Von Heijne matrix
 - (D) OTHER INFORMATION: score 7.2

seq ILFSLSFLLVIIT/FP

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 208:

Met Asp Ser Arg Val Ser Ser Pro Glu Lys Gln Asp Lys Glu Asn Phe
-40 -35 -30

Val Gly Val Asn Asn Lys Arg Leu Gly Val Cys Gly Trp Ile Leu Phe
-25
-15

Ser Leu Ser Phe Leu Leu Val Ile Ile Thr Phe Pro Ile Ser Ile Trp
-10 -5 1 5

Met Cys Leu Lys Ile Ile Lys Xaa Tyr Glu Arg Xaa Val Val Phe Arg
10 15 20

Leu Gly Arg His Gly 25

- (2) INFORMATION FOR SEQ ID NO: 209:
 - (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 18 amino acids
 - (B) TYPE: AMINO ACID
 - (D) TOPOLOGY: LINEAR
 - (ii) MOLECULE TYPE: PROTEIN
 - (vi) ORIGINAL SOURCE:
 - (A) ORGANISM: Homo Sapiens
 - (F) TISSUE TYPE: Thyroid
 - (ix) FEATURE:
 - (A) NAME/KEY: sig_peptide
 - (B) LOCATION: -16..-1
 - (C) IDENTIFICATION METHOD: Von Heijne matrix
 - (D) OTHER INFORMATION: score 7

seq LFCVVLCLSPTSY/CY

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 209:

Met Cys Ile Leu Phe Cys Val Val Leu Cys Leu Ser Pro Thr Ser Tyr
-15 -5

Cys Tyr

- (2) INFORMATION FOR SEQ ID NO: 210:
 - (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 28 amino acids
 - (B) TYPE: AMINO ACID
 - (D) TOPOLOGY: LINEAR
 - (ii) MOLECULE TYPE: PROTEIN
 - (vi) ORIGINAL SOURCE:
 - (A) ORGANISM: Homo Sapiens
 - (F) TISSUE TYPE: Lung (cells)
 - (ix) FEATURE:
 - (A) NAME/KEY: sig_peptide
 - (B) LOCATION: -19..-1
 - (C) IDENTIFICATION METHOD: Von Heijne matrix
 - (D) OTHER INFORMATION: score 7 seq ETLLCLGSSCCQC/RI
 - (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 210:

Met His Arg Gly Asp Ile Glu Thr Leu Leu Cys Leu Gly Ser Ser Cys -15 -10 -5

Cys Gln Cys Arg Ile Phe Ser Phe Phe Phe Phe Phe 5

- (2) INFORMATION FOR SEQ ID NO: 211:
 - (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 27 amino acids
 - (B) TYPE: AMINO ACID
 - (D) TOPOLOGY: LINEAR
 - (ii) MOLECULE TYPE: PROTEIN
 - (vi) ORIGINAL SOURCE:
 - (A) ORGANISM: Homo Sapiens
 - (F) TISSUE TYPE: Lung
 - (ix) FEATURE:
 - (A) NAME/KEY: sig_peptide
 - (B) LOCATION: -19..-1
 - (C) IDENTIFICATION METHOD: Von Heijne matrix
 - (D) OTHER INFORMATION: score 7

184

seq AGLSSCLLPLCWL/ER

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 211:

Met Phe Leu Lys Ser Gly Ala Gly Leu Ser Ser Cys Leu Leu Pro Leu
-15 -10 -5

Cys Trp Leu Glu Arg Lys Asp His Gly Arg Arg

- (2) INFORMATION FOR SEQ ID NO: 212:
 - (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 120 amino acids
 - (B) TYPE: AMINO ACID
 - (D) TOPOLOGY: LINEAR
 - (ii) MOLECULE TYPE: PROTEIN
 - (vi) ORIGINAL SOURCE:
 - (A) ORGANISM: Homo Sapiens
 - (F) TISSUE TYPE: Lung (cells)
 - (ix) FEATURE:
 - (A) NAME/KEY: sig_peptide
 - (B) LOCATION: -38..-1
 - (C) IDENTIFICATION METHOD: Von Heijne matrix
 - (D) OTHER INFORMATION: score 7 seq LVMVWLGLLPLFS/GP
 - (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 212:
- Met Ala Asn Ala Ile Ile Lys Lys Pro Cys Ala Met Pro Ala Gln Pro
 -35 -30 -25
- His Thr Gly Asn Leu Leu Trp Pro Pro Leu Val Met Val Trp Leu Gly
 -20 -15 -10
- Leu Leu Pro Leu Phe Ser Gly Pro His Leu Gln Ala Val Gln His Leu
 -5 1 5 10
- Ala Leu Ala Tyr Leu Pro Leu Asn Ser Val Val Leu Ala His Asn Ser 15 20 25
- Pro Ala Ile Leu Asn Val Trp Leu Thr Leu Arg Cys Pro Leu Pro Tyr 30 35 40
- Arg Ile Cys Leu Trp Pro Phe Glu His Ala Phe Pro Ser Ile Arg Asn 45 50 55
- Thr His Ser Cys Leu Ser Ser Ser Cys Cys Cys Pro Ala Ser Ala Pro 60 65 70
- Leu Leu Val Asp Tyr Leu Phe Phe 75 80

- (2) INFORMATION FOR SEQ ID NO: 213:
 - (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 130 amino acids
 - (B) TYPE: AMINO ACID
 - (D) TOPOLOGY: LINEAR
 - (ii) MOLECULE TYPE: PROTEIN
 - (vi) ORIGINAL SOURCE:
 - (A) ORGANISM: Homo Sapiens
 - (F) TISSUE TYPE: Lung
 - (ix) FEATURE:
 - (A) NAME/KEY: sig_peptide
 - (B) LOCATION: -19..-1
 - (C) IDENTIFICATION METHOD: Von Heijne matrix
 - (D) OTHER INFORMATION: score 6.9

seq LQPLLLLPLLNV/EP

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 213:

Met Ser Pro Pro Pro Leu Leu Gln Pro Leu Leu Leu Leu Leu Pro Leu
-15
-10
-5

Leu Asn Val Glu Pro Ser Gly Ala Thr Leu Ile Arg Ile Pro Leu His $1 \hspace{1cm} 5 \hspace{1cm} 10$

Arg Val Gln Pro Gly Arg Arg Ile Leu Asn Leu Leu Arg Gly Trp Xaa 15 20 25

Glu Pro Ala Glu Leu Pro Lys Leu Gly Ala Pro Ser Pro Gly Asp Lys 30 35 40 45

Pro Ile Phe Val Pro Leu Ser Asn Tyr Arg Asp Val Gln Tyr Phe Gly
50 55

Glu Ile Gly Leu Gly Thr Pro Pro Gln Asn Phe Thr Val Ala Phe Asp
65 70 75

Thr Gly Ser Ser Asn Leu Trp Val Pro Ser Arg Arg Cys His Phe Phe 80 85 90

Ser Val Pro Cys Trp Leu His Thr Asp Leu Ile Pro Lys Pro Leu Ala 95 100 105

Pro Ser 110

- (2) INFORMATION FOR SEQ ID NO: 214:
 - (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 44 amino acids
 - (B) TYPE: AMINO ACID
 - (D) TOPOLOGY: LINEAR
 - (ii) MOLECULE TYPE: PROTEIN

PCT/IB98/01233 WO 99/06439

186

(vi) ORIGINAL SOURCE:

- (A) ORGANISM: Homo Sapiens
- (D) DEVELOPMENTAL STAGE: Fetal
- (F) TISSUE TYPE: liver
- (ix) FEATURE:
 - (A) NAME/KEY: sig_peptide
 - (B) LOCATION: -42..-1
 - (C) IDENTIFICATION METHOD: Von Heijne matrix
 - (D) OTHER INFORMATION: score 6.5

seq ILIGLFSLTGLVA/GN

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 214:

Met Ile Pro Ile Tyr Gln Asn Lys Ser Gln Thr Asp Ser His Cys Ser -35

Leu Ser His Lys Gly Leu Ala Phe Leu Lys Val Trp Leu Ile Leu Ile -15

Gly Leu Phe Ser Leu Thr Gly Leu Val Ala Gly Asn

- (2) INFORMATION FOR SEQ ID NO: 215:
 - (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 40 amino acids
 - (B) TYPE: AMINO ACID
 - (D) TOPOLOGY: LINEAR
 - (ii) MOLECULE TYPE: PROTEIN
 - (vi) ORIGINAL SOURCE:
 - (A) ORGANISM: Homo Sapiens
 - (F) TISSUE TYPE: Lung (cells)
 - (ix) FEATURE:
 - (A) NAME/KEY: sig_peptide
 - (B) LOCATION: -20..-1
 - (C) IDENTIFICATION METHOD: Von Heijne matrix
 - (D) OTHER INFORMATION: score 6.5 seq LLSGSTCPGPCSC/GS
 - (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 215:

Met Ala Leu Pro Gly Ile His Leu Leu Ser Gly Ser Thr Cys Pro Gly

Pro Cys Ser Cys Gly Ser Leu Arg Ser Pro Pro Gly Pro Val Thr Asp

Lys Pro Leu Pro Leu Pro Pro Gln 15

- (2) INFORMATION FOR SEQ ID NO: 216:
 - (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 62 amino acids
 - (B) TYPE: AMINO ACID
 - (D) TOPOLOGY: LINEAR
 - (ii) MOLECULE TYPE: PROTEIN
 - (vi) ORIGINAL SOURCE:
 - (A) ORGANISM: Homo Sapiens
 - (F) TISSUE TYPE: Colon
 - (ix) FEATURE:
 - (A) NAME/KEY: sig_peptide
 - (B) LOCATION: -41..-1
 - (C) IDENTIFICATION METHOD: Von Heijne matrix
 - (D) OTHER INFORMATION: score 6.4

seq IALIPLFSTXAFA/IX

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 216:

Met Pro Ser Glu Thr Leu Trp Glu Ile Ala Lys Ala Glu Val Glu Lys
-40 -35 -30

Arg Gly Ile Asn Gly Xaa Xaa Gly Asp Gly Ala Glu Ile Ala Leu Ile
-25 -20 -15 -10

Pro Leu Phe Ser Thr Xaa Ala Phe Ala Ile Xaa Gln Ile Val Ser Leu
-5 1 5

Gly Ile Val Asp Gly Ser Xaa Pro Pro Xaa Ser Arg Thr Pro 10 15 20

- (2) INFORMATION FOR SEQ ID NO: 217:
 - (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 58 amino acids
 - (B) TYPE: AMINO ACID
 - (D) TOPOLOGY: LINEAR
 - (ii) MOLECULE TYPE: PROTEIN
 - (vi) ORIGINAL SOURCE:
 - (A) ORGANISM: Homo Sapiens
 - (F) TISSUE TYPE: Lung (cells)
 - (ix) FEATURE:
 - (A) NAME/KEY: sig_peptide
 - (B) LOCATION: -37..-1
 - (C) IDENTIFICATION METHOD: Von Heijne matrix
 - (D) OTHER INFORMATION: score 6.4

seq LCMSLTFLALSTL/RF

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 217:

Met Glu Trp Leu Arg Pro Ser Gln Ile Ser Phe Tyr Pro Gly Tyr Ser

188

-35 **-**30 **-**25

Lys Glu Arg Leu Arg Leu Val Leu Cys Met Ser Leu Thr Phe Leu
-20 -15 -10

Ala Leu Ser Thr Leu Arg Phe Leu Thr Gln Arg Val Gln Met Gln Ala -5 10

Gly Cys Pro Leu Arg Ser Pro Arg Leu Trp 15 20

- (2) INFORMATION FOR SEQ ID NO: 218:
 - (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 41 amino acids
 - (B) TYPE: AMINO ACID
 - (D) TOPOLOGY: LINEAR
 - (ii) MOLECULE TYPE: PROTEIN
 - (vi) ORIGINAL SOURCE:
 - (A) ORGANISM: Homo Sapiens
 - (F) TISSUE TYPE: Liver
 - (ix) FEATURE:
 - (A) NAME/KEY: sig_peptide
 - (B) LOCATION: -18..-1
 - (C) IDENTIFICATION METHOD: Von Heijne matrix
 - (D) OTHER INFORMATION: score 6.3

seq HLTLLALLSVNTG/KE

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 218:

Met Lys Ala Ile Ile His Leu Thr Leu Leu Ala Leu Leu Ser Val Asn
-15 -10 -5

Thr Gly Lys Glu Tyr Phe Tyr Ile Leu Ile Leu Pro Ile Met Tyr Val 1 5 10

Val Phe Glu Val Glu Ser Ala Gly Gln 15 20

- (2) INFORMATION FOR SEQ ID NO: 219:
 - (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 45 amino acids
 - (B) TYPE: AMINO ACID
 - (D) TOPOLOGY: LINEAR
 - (ii) MOLECULE TYPE: PROTEIN
 - (vi) ORIGINAL SOURCE:
 - (A) ORGANISM: Homo Sapiens
 - (F) TISSUE TYPE: Liver
 - (ix) FEATURE:

189

- (A) NAME/KEY: sig_peptide
 - (B) LOCATION: -29..-1
 - (C) IDENTIFICATION METHOD: Von Heijne matrix
 - (D) OTHER INFORMATION: score 6.3 seq SLPLALTLSLSTS/LH
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 219:

Met Asp Val Ser Ala Ser Lys Pro Val Ala Glu Ser Trp Ser Pro Gly
-25 -20 -15

Ser Leu Pro Leu Ala Leu Thr Leu Ser Leu Ser Thr Ser Leu His Asp
-10 -5 1

Ser Trp Lys Glu Pro Ile Pro Asn Leu His Gln Pro Ala
5 10 15

- (2) INFORMATION FOR SEQ ID NO: 220:
 - (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 39 amino acids
 - (B) TYPE: AMINO ACID
 - (D) TOPOLOGY: LINEAR
 - (ii) MOLECULE TYPE: PROTEIN
 - (vi) ORIGINAL SOURCE:
 - (A) ORGANISM: Homo Sapiens
 - (F) TISSUE TYPE: Liver
 - (ix) FEATURE:
 - (A) NAME/KEY: sig_peptide
 - (B) LOCATION: -28..-1
 - (C) IDENTIFICATION METHOD: Von Heijne matrix
 - (D) OTHER INFORMATION: score 6.2 seq IQTALLGLPXAWA/SS
 - (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 220:

Met Gly Val Arg Val Gly Val Ser Leu Arg Ala Trp Cys Val Phe Ile -25 -20 -15

Gln Thr Ala Leu Leu Gly Leu Pro Xaa Ala Trp Ala Ser Ser Gly Val
-10 -5 1

Val Ser Ser Thr Gly Pro Gly 5 10

- (2) INFORMATION FOR SEQ ID NO: 221:
 - (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 29 amino acids
 - (B) TYPE: AMINO ACID
 - (D) TOPOLOGY: LINEAR

190

- (ii) MOLECULE TYPE: PROTEIN
- (vi) ORIGINAL SOURCE:
 - (A) ORGANISM: Homo Sapiens
 - (F) TISSUE TYPE: Lung (cells)
- (ix) FEATURE:
 - (A) NAME/KEY: sig_peptide
 - (B) LOCATION: -22..-1
 - (C) IDENTIFICATION METHOD: Von Heijne matrix
 - (D) OTHER INFORMATION: score 6.1

seq FFLLLCIPFLTLL/LY

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 221:

Met Ile Ile Ser Ile Ile Pro Arg Ser Phe Phe Leu Leu Cys Ile
-20 -15 -10

Pro Phe Leu Thr Leu Leu Leu Tyr Thr Tyr Pro Pro Arg
-5 5

- (2) INFORMATION FOR SEQ ID NO: 222:
 - (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 54 amino acids
 - (B) TYPE: AMINO ACID
 - (D) TOPOLOGY: LINEAR
 - (ii) MOLECULE TYPE: PROTEIN
 - (vi) ORIGINAL SOURCE:
 - (A) ORGANISM: Homo Sapiens
 - (F) TISSUE TYPE: Thyroid
 - (ix) FEATURE:
 - (A) NAME/KEY: sig_peptide
 - (B) LOCATION: -29..-1
 - (C) IDENTIFICATION METHOD: Von Heijne matrix
 - (D) OTHER INFORMATION: score 5.9

seq WTLVLMSPEWALL/QY

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 222:

Met Thr Met Gln Arg Ser Arg Ser Ser Ser Trp Thr Ser Cys Asn Ser
-25 -20 -15

Trp Thr Leu Val Leu Met Ser Pro Glu Trp Ala Leu Leu Gln Tyr Gly -10 -5 1

Ser Thr Val Lys Asn Glu Phe Ser Xaa Lys Thr Phe Lys Arg Lys Ser 5 10 15

Glu Val Glu Arg Ala Val 20 25

191

(2) INFORMATION FOR SEQ ID NO: 223:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 30 amino acids
 - (B) TYPE: AMINO ACID
 - (D) TOPOLOGY: LINEAR
- (ii) MOLECULE TYPE: PROTEIN
- (vi) ORIGINAL SOURCE:
 - (A) ORGANISM: Homo Sapiens
 - (F) TISSUE TYPE: Lung (cells)
- (ix) FEATURE:
 - (A) NAME/KEY: sig_peptide
 - (B) LOCATION: -23..-1
 - (C) IDENTIFICATION METHOD: Von Heijne matrix
 - (D) OTHER INFORMATION: score 5.8 seq LCSLMASISPTLT/AV
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 223:

Met Ile Thr Leu Pro Gln Thr Ser Ser Leu Leu Cys Ser Leu Met Ala
-20 -15 -10

Ser Ile Ser Pro Thr Leu Thr Ala Val Ile Leu Trp Pro Pro
-5 1 5

- (2) INFORMATION FOR SEQ ID NO: 224:
 - (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 78 amino acids
 - (B) TYPE: AMINO ACID
 - (D) TOPOLOGY: LINEAR
 - (ii) MOLECULE TYPE: PROTEIN
 - (vi) ORIGINAL SOURCE:
 - (A) ORGANISM: Homo Sapiens
 - (F) TISSUE TYPE: Liver
 - (ix) FEATURE:
 - (A) NAME/KEY: sig_peptide
 - (B) LOCATION: -63..-1
 - (C) IDENTIFICATION METHOD: Von Heijne matrix
 - (D) OTHER INFORMATION: score 5.7

seq PLVTHGLLLQAWS/RR

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 224:

Met Leu Arg Thr Cys Tyr Val Leu Cys Ser Gln Ala Gly Pro Pro Ser
-60 -55 -50

Arg Gly Trp Gln Ser Leu Ser Phe Asp Gly Gly Ala Phe His Leu Lys
-45
-40
-35

Gly Thr Gly Glu Leu Thr Arg Ala Leu Leu Val Leu Arg Leu Cys Ala

192

-30 -25 -20

Trp Pro Pro Leu Val Thr His Gly Leu Leu Gln Ala Trp Ser Arg
-15 -5 1

Arg Leu Leu Gly Ser Arg Leu Ser Xaa Ala Phe Leu Arg Ala 5 10 15

- (2) INFORMATION FOR SEQ ID NO: 225:
 - (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 41 amino acids
 - (B) TYPE: AMINO ACID
 - (D) TOPOLOGY: LINEAR
 - (ii) MOLECULE TYPE: PROTEIN
 - (vi) ORIGINAL SOURCE:
 - (A) ORGANISM: Homo Sapiens
 - (F) TISSUE TYPE: Lung (cells)
 - (ix) FEATURE:
 - (A) NAME/KEY: sig_peptide
 - (B) LOCATION: -30..-1
 - (C) IDENTIFICATION METHOD: Von Heijne matrix
 - (D) OTHER INFORMATION: score 5.7

seq LGVGCHFFHLALG/RF

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 225:

Met Ile Cys Ser Pro Phe Ser Gly Phe Ala Pro Cys Gln Ala Leu Gly
-30 -25 -20 -15

Thr Leu Gly Val Gly Cys His Phe Phe His Leu Ala Leu Gly Arg Phe -10 -5 1

Leu Leu Ser Leu Ser Asn Asn Ile Tyr
5 10

- (2) INFORMATION FOR SEQ ID NO: 226:
 - (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 150 amino acids
 - (B) TYPE: AMINO ACID
 - (D) TOPOLOGY: LINEAR
 - (ii) MOLECULE TYPE: PROTEIN
 - (vi) ORIGINAL SOURCE:
 - (A) ORGANISM: Homo Sapiens
 - (F) TISSUE TYPE: Pancreas
 - (ix) FEATURE:
 - (A) NAME/KEY: sig_peptide
 - (B) LOCATION: -98..-1
 - (C) IDENTIFICATION METHOD: Von Heijne matrix

193

(D) OTHER INFORMATION: score 5.6

seq LLLLRGADRVLQA/HI

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 226:

Met Cys Asn Pro Glu Glu Ala Ala Leu Xaa Gly Leu Glu Glu Val Phe -95 -90 -85

Ser Ala Thr Leu Ala His Val Asn Ser Leu Val Leu Gln Pro Leu Leu
-80 -75 -70

Pro Ala Ala Pro Asp Pro Ser Asp Pro Trp Gly Arg Glu Cys Leu Arg
-65 -60 -55

Leu Leu Gln Gln Leu His Lys Ser Ser Gln Gln Leu Trp Glu Val Thr
-50 -45 -40 -35

Glu Glu Ser Leu His Ser Leu Gln Glu Arg Leu Arg Tyr Pro Asp Ser
-30 -25 -20

Thr Gly Leu Glu Ser Leu Leu Leu Leu Arg Gly Ala Asp Arg Val Leu
-15 -10 -5

Gln Ala His Ile Glu Tyr Ile Glu Ser Tyr Thr Ser Cys Met Val Val 1 5

Gin Ala Phe Gin Lys Xaa Ala Lys Arg Arg Ser Glu Phe Trp Arg Gly 15 20 25 30

Gln Arg Xaa Ala Leu Arg Gln Leu Leu Ser Gly Val Ser Ser Glu Gly 35 40 45

Ser Val Gly Ala Ser Leu 50

(2) INFORMATION FOR SEQ ID NO: 227:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 70 amino acids
 - (B) TYPE: AMINO ACID
 - (D) TOPOLOGY: LINEAR
- (ii) MOLECULE TYPE: PROTEIN
- (vi) ORIGINAL SOURCE:
 - (A) ORGANISM: Homo Sapiens
 - (F) TISSUE TYPE: Thyroid
- (ix) FEATURE:
 - (A) NAME/KEY: sig_peptide
 - (B) LOCATION: -45..-1
 - (C) IDENTIFICATION METHOD: Von Heijne matrix
 - (D) OTHER INFORMATION: score 5.6

seq LLVIHWVMCPSLS/QS

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 227:

194

Met Asp Lys Leu Ile Pro Ser Leu Ser Ser Gln Glu Asn Arg Lys Ala
-45 -40 -35 -30

Ser His Thr Leu His Lys Ala Arg Asn Lys Gln His Cys Gly Gly Phe
-25 -20 -15

Leu Leu Val Ile His Trp Val Met Cys Pro Ser Leu Ser Gln Ser Ala
-10 -5 1

Val Arg Arg Met Lys Tyr Ser Asn Trp Pro Val Leu Gly His Val Pro
5 10 15

Val Pro Gly Cys His Cys 20 25

- (2) INFORMATION FOR SEQ ID NO: 228:
 - (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 35 amino acids
 - (B) TYPE: AMINO ACID
 - (D) TOPOLOGY: LINEAR
 - (ii) MOLECULE TYPE: PROTEIN
 - (vi) ORIGINAL SOURCE:
 - (A) ORGANISM: Homo Sapiens
 - (F) TISSUE TYPE: Lung (cells)
 - (ix) FEATURE:
 - (A) NAME/KEY: sig_peptide
 - (B) LOCATION: -17..-1
 - (C) IDENTIFICATION METHOD: Von Heijne matrix
 - (D) OTHER INFORMATION: score 5.5 seq LPVVLASPPVGHG/LP
 - (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 228:

Met Ser Xaa Leu Leu Pro Val Val Leu Ala Ser Pro Pro Val Gly His
-15 -10 -5

Gly Leu Pro Ser Pro Val Pro Leu Leu Gln Asp Pro Cys Pro Leu Pro 1 5 10 15

Ala Val Gly

- (2) INFORMATION FOR SEQ ID NO: 229:
 - (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 25 amino acids
 - (B) TYPE: AMINO ACID
 - (D) TOPOLOGY: LINEAR
 - (ii) MOLECULE TYPE: PROTEIN
 - (vi) ORIGINAL SOURCE:
 - (A) ORGANISM: Homo Sapiens

195

(F) TISSUE TYPE: Lung (cells)

- (ix) FEATURE:
 - (A) NAME/KEY: sig_peptide
 - (B) LOCATION: -13..-1
 - (C) IDENTIFICATION METHOD: Von Heijne matrix
 - (D) OTHER INFORMATION: score 5.5

seq MVLLTMIARVADG/LP

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 229:

Met Val Leu Thr Met Ile Ala Arg Val Ala Asp Gly Leu Pro Leu
-10 -5 1

Ala Ala Ser Met Gln Glu Asp Glu Glu
5 10

- (2) INFORMATION FOR SEQ ID NO: 230:
 - (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 59 amino acids
 - (B) TYPE: AMINO ACID
 - (D) TOPOLOGY: LINEAR
 - (ii) MOLECULE TYPE: PROTEIN
 - (vi) ORIGINAL SOURCE:
 - (A) ORGANISM: Homo Sapiens
 - (F) TISSUE TYPE: Large intestine
 - (ix) FEATURE:
 - (A) NAME/KEY: sig_peptide
 - (B) LOCATION: -53..-1
 - (C) IDENTIFICATION METHOD: Von Heijne matrix
 - (D) OTHER INFORMATION: score 5.5

seq LLELLFVPLLCFL/SK

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 230:

Met Phe His Ile Ala Phe Ser Glu Ala Leu Pro Val Asp Ile Phe Lys
-50 -45 -40

Thr Gln Pro Asn Cys His Glu Ala Phe Ser Met Lys Ala Ile His Ile
-35 -30 -25

Thr Arg Ile Arg Ser Gly Leu Cys Leu Leu Glu Leu Leu Phe Val Pro
-20 -15 -10

Leu Leu Cys Phe Leu Ser Lys Lys Trp Pro Trp -5 5

- (2) INFORMATION FOR SEQ ID NO: 231:
 - (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 38 amino acids

196

(B) TYPE: AMINO ACID

- (D) TOPOLOGY: LINEAR
- (ii) MOLECULE TYPE: PROTEIN
- (vi) ORIGINAL SOURCE:
 - (A) ORGANISM: Homo Sapiens
 - (F) TISSUE TYPE: Pancreas
- (ix) FEATURE:
 - (A) NAME/KEY: sig_peptide
 - (B) LOCATION: -25..-1
 - (C) IDENTIFICATION METHOD: Von Heijne matrix
 - (D) OTHER INFORMATION: score 5.4

seq SLLTETVLPLAFP/GP

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 231:

Met Met His Cys Thr Pro Ser Gly Ser Ala Ala Val Ser Leu Leu Thr
-25 -10 -15

Glu Thr Val Leu Pro Leu Ala Phe Pro Gly Pro Pro Trp Leu Gly Thr
-5 1 5

Ser Phe Asn Arg Xaa Leu 10

- (2) INFORMATION FOR SEQ ID NO: 232:
 - (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 74 amino acids
 - (B) TYPE: AMINO ACID
 - (D) TOPOLOGY: LINEAR
 - (ii) MOLECULE TYPE: PROTEIN
 - (vi) ORIGINAL SOURCE:
 - (A) ORGANISM: Homo Sapiens
 - (F) TISSUE TYPE: Lung
 - (ix) FEATURE:
 - (A) NAME/KEY: sig_peptide
 - (B) LOCATION: -28..-1
 - (C) IDENTIFICATION METHOD: Von Heijne matrix
 - (D) OTHER INFORMATION: score 5.3

seq RISCAFSLASSTA/RQ

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 232:

Met Thr Arg Pro Phe Trp Ala Ser Cys Ser Thr Trp Ala Thr Ser Arg -25 -20 -15

Ile Ala Cys Cys Ala Thr His Arg Thr Ala Trp Ala Ser Arg Pro Gly
5 10 15 20

Pro Arg Arg Pro Trp Cys Cys Arg Tyr Ser Lys Pro Leu Thr Trp
25 30 35

Pro Val Arg Met Met Arg Arg Glu Gly Leu
40 45

- (2) INFORMATION FOR SEQ ID NO: 233:
 - (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 87 amino acids
 - (B) TYPE: AMINO ACID
 - (D) TOPOLOGY: LINEAR
 - (ii) MOLECULE TYPE: PROTEIN
 - (vi) ORIGINAL SOURCE:
 - (A) ORGANISM: Homo Sapiens
 - (F) TISSUE TYPE: Lung (cells)
 - (ix) FEATURE:
 - (A) NAME/KEY: sig_peptide
 - (B) LOCATION: -26..-1
 - (C) IDENTIFICATION METHOD: Von Heijne matrix
 - (D) OTHER INFORMATION: score 5.2

seq SCCLIQWPELSFS/NT

- (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 233:
- Met Val Thr His Leu Ile Arg Gly Val Val Leu Gln Gly Ser Cys Cys
 -25 -20 -15
- Leu Ile Gln Trp Pro Glu Leu Ser Phe Ser Asn Thr Asn Gly Val Cys
 -10 -5 1 5
- Pro Ile Tyr Pro Pro Pro Ser Ile Xaa Xaa Leu Arg Met Ser Ser Cys
 10 15 20
- Ser Pro Leu Thr Val Ser Leu Cys Pro Cys Tyr Val Glu Cys Ala Ser 25 30 35
- Thr Pro Gly Pro Leu Cys Leu Leu Phe Ser Trp Pro Arg Asn Thr Ser

Pro Asn Met Pro Ser Gly Tyr
55 60

- (2) INFORMATION FOR SEQ ID NO: 234:
 - (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 65 amino acids
 - (B) TYPE: AMINO ACID
 - (D) TOPOLOGY: LINEAR
 - (ii) MOLECULE TYPE: PROTEIN

198

- (vi) ORIGINAL SOURCE:
 - (A) ORGANISM: Homo Sapiens
 - (F) TISSUE TYPE: Lung (cells)
- (ix) FEATURE:
 - (A) NAME/KEY: sig_peptide
 - (B) LOCATION: -19..-1
 - (C) IDENTIFICATION METHOD: Von Heijne matrix
 - (D) OTHER INFORMATION: score 5.2

seq LLVAFRVFLGLFS/LP

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 234:

Met Tyr Met Trp Ser Lys Leu Leu Val Ala Phe Arg Val Phe Leu Gly
-15 -10

Leu Phe Ser Leu Pro Ser Asn His Asn Thr Tyr Cys Pro Phe Gln Pro 1 5 10

Trp Gly Ile Pro Cys Ser Leu Arg Ile Gly Gly Leu Leu His Leu Gln
15 20 25

Cys Pro Leu Pro Pro Ser Leu His Pro Leu Pro Ser Leu Leu Thr Ser 30 40 45

Arg

- (2) INFORMATION FOR SEQ ID NO: 235:
 - (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 45 amino acids
 - (B) TYPE: AMINO ACID
 - (D) TOPOLOGY: LINEAR
 - (ii) MOLECULE TYPE: PROTEIN
 - (vi) ORIGINAL SOURCE:
 - (A) ORGANISM: Homo Sapiens
 - (F) TISSUE TYPE: Lung (cells)
 - (ix) FEATURE:
 - (A) NAME/KEY: sig_peptide
 - (B) LOCATION: -23..-1
 - (C) IDENTIFICATION METHOD: Von Heijne matrix
 - (D) OTHER INFORMATION: score 5.2

seq FLFGLYSFRAVDS/SR

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 235:

Met Ser Ser Arg Asn Cys Phe Phe Pro Ser Phe Leu Phe Gly Leu Tyr

Ser Phe Arg Ala Val Asp Ser Ser Arg Ile Lys Leu Ser Leu Leu Thr
-5 1 5

Lys Glu Glu Glu Thr Pro Ser Ala Tyr Tyr Arg Ser Leu 10 20

PCT/IB98/01233

- (2) INFORMATION FOR SEQ ID NO: 236:
 - (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 36 amino acids
 - (B) TYPE: AMINO ACID
 - (D) TOPOLOGY: LINEAR
 - (ii) MOLECULE TYPE: PROTEIN
 - (vi) ORIGINAL SOURCE:
 - (A) ORGANISM: Homo Sapiens
 - (F) TISSUE TYPE: Lung (cells)
 - (ix) FEATURE:
 - (A) NAME/KEY: sig_peptide
 - (B) LOCATION: -18..-1
 - (C) IDENTIFICATION METHOD: Von Heijne matrix
 - (D) OTHER INFORMATION: score 5.1

seq CLYLHVYVLTCSG/CN

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 236:

Met Tyr Met Asn Thr Cys Leu Tyr Leu His Val Tyr Val Leu Thr Cys
-15 -10 -5

Ser Gly Cys Asn Val Asp Met Cys Ser Arg Leu Phe Leu Ser Thr Lys

1 10

Leu Lys Ala Arg

- (2) INFORMATION FOR SEQ ID NO: 237:
 - (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 56 amino acids
 - (B) TYPE: AMINO ACID
 - (D) TOPOLOGY: LINEAR
 - (ii) MOLECULE TYPE: PROTEIN
 - (vi) ORIGINAL SOURCE:
 - (A) ORGANISM: Homo Sapiens
 - (F) TISSUE TYPE: Lung (cells)
 - (ix) FEATURE:
 - (A) NAME/KEY: sig_peptide
 - (B) LOCATION: -49..-1
 - (C) IDENTIFICATION METHOD: Von Heijne matrix
 - (D) OTHER INFORMATION: score 5.1

seq VALSASLPQCSLG/LL

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 237:

Met Ser Cys Arg Gln Pro Thr Pro Thr Gln Cys Ser Leu Leu Pro Asn

200

-45 -40 -35

Asp Asn Arg Val Ser Thr Arg Gly Gly Asp Ser Ala Gly Arg His Arg
-30 -25 -20

Gln Val Pro Gln Val Ala Leu Ser Ala Ser Leu Pro Gln Cys Ser Leu
-15 -10 -5

Gly Leu Leu Ile Asn Pro Arg Leu
1 5

- (2) INFORMATION FOR SEQ ID NO: 238:
 - (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 25 amino acids
 - (B) TYPE: AMINO ACID
 - (D) TOPOLOGY: LINEAR
 - (ii) MOLECULE TYPE: PROTEIN
 - (vi) ORIGINAL SOURCE:
 - (A) ORGANISM: Homo Sapiens
 - (F) TISSUE TYPE: Thyroid
 - (ix) FEATURE:
 - (A) NAME/KEY: sig peptide
 - (B) LOCATION: -20..-1
 - (C) IDENTIFICATION METHOD: Von Heijne matrix
 - (D) OTHER INFORMATION: score 5.1 seq PTAGVVVLQGSRA/SV
 - (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 238:

Met Ile Thr Gly Cys Thr Lys Pro Thr Ala Gly Val Val Leu Gln -20 -15 -10 -5

Gly Ser Arg Ala Ser Val Arg Gln Arg

- (2) INFORMATION FOR SEQ ID NO: 239:
 - (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 17 amino acids
 - (B) TYPE: AMINO ACID
 - (D) TOPOLOGY: LINEAR
 - (ii) MOLECULE TYPE: PROTEIN
 - (vi) ORIGINAL SOURCE:
 - (A) ORGANISM: Homo Sapiens
 - (F) TISSUE TYPE: Lung (cells)
 - (ix) FEATURE:
 - (A) NAME/KEY: sig_peptide
 - (B) LOCATION: -14..-1
 - (C) IDENTIFICATION METHOD: Von Heijne matrix

201

(D) OTHER INFORMATION: score 5 seq GLDLILSFSSSSP/GP

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 239:

Met Gly Leu Asp Leu Ile Leu Ser Phe Ser Ser Ser Pro Gly Pro
-10 -5

Gly

- (2) INFORMATION FOR SEQ ID NO: 240:
 - (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 69 amino acids
 - (B) TYPE: AMINO ACID
 - (D) TOPOLOGY: LINEAR
 - (ii) MOLECULE TYPE: PROTEIN
 - (vi) ORIGINAL SOURCE:
 - (A) ORGANISM: Homo Sapiens
 - (F) TISSUE TYPE: Colon
 - (ix) FEATURE:
 - (A) NAME/KEY: sig_peptide
 - (B) LOCATION: -62..-1
 - (C) IDENTIFICATION METHOD: Von Heijne matrix
 - (D) OTHER INFORMATION: score 4.9 seq LDRLCALTSLCSP/GP
 - (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 240:

Met Arg Glu Asp Asn Glu His Glu Arg Asn Val Pro Ser Gly Val Glu
-60 -55 -50

Asn Val Lys Glu Glu Gly Gly Asp Glu Asp Leu Ser Trp Gly Asp Glu
-45 -35

Gly Cys Gln Val Leu Arg His Arg Leu Arg Val Cys Arg Lys Val Gly
-30
-25
-20
-15

Leu Leu Asp Arg Leu Cys Ala Leu Thr Ser Leu Cys Ser Pro Gly Pro
-10 -5 1

Leu Pro Ala Thr Leu

- (2) INFORMATION FOR SEQ ID NO: 241:
 - (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 29 amino acids
 - (B) TYPE: AMINO ACID
 - (D) TOPOLOGY: LINEAR
 - (ii) MOLECULE TYPE: PROTEIN

- (vi) ORIGINAL SOURCE:
 - (A) ORGANISM: Homo Sapiens
 - (F) TISSUE TYPE: Colon
- (ix) FEATURE:
 - (A) NAME/KEY: sig peptide
 - (B) LOCATION: -18..-1
 - (C) IDENTIFICATION METHOD: Von Heijne matrix
 - (D) OTHER INFORMATION: score 4.9

seq GAVVSSWAXCSLG/XP

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 241:

Met Gly Lys Arg Ala Gly Ala Val Val Ser Ser Trp Ala Xaa Cys Ser
-15 -10 -5

Leu Gly Xaa Pro Gly Ile Gln Arg Ser Ser Arg Leu Thr
1 5 10

- (2) INFORMATION FOR SEQ ID NO: 242:
 - (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 82 amino acids
 - (B) TYPE: AMINO ACID
 - (D) TOPOLOGY: LINEAR
 - (ii) MOLECULE TYPE: PROTEIN
 - (vi) ORIGINAL SOURCE:
 - (A) ORGANISM: Homo Sapiens
 - (F) TISSUE TYPE: Lung (cells)
 - (ix) FEATURE:
 - (A) NAME/KEY: sig_peptide
 - (B) LOCATION: -21..-1
 - (C) IDENTIFICATION METHOD: Von Heijne matrix
 - (D) OTHER INFORMATION: score 4.9

seq WLLSDILGQGATA/NV

- (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 242:
- Met Gln Ser Thr Ser Asn His Leu Trp Leu Leu Ser Asp Ile Leu Gly
 -20 -15 -10
- Gln Gly Ala Thr Ala Asn Val Phe Arg Gly Arg His Lys Lys Thr Gly -5 10
- Asp Leu Phe Ala Ile Lys Val Phe Asn Asn Ile Ser Phe Leu Arg Pro 15 20 25
- Val Asp Val Gln Met Arg Glu Phe Glu Val Leu Lys Lys Leu Asn His 30 35 40
- Lys Asn Ile Val Lys Leu Phe Ala Ile Glu Glu Glu Thr Thr Arg
 45 50 55

203

Arg Arg 60

- (2) INFORMATION FOR SEQ ID NO: 243:
 - (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 67 amino acids
 - (B) TYPE: AMINO ACID
 - (D) TOPOLOGY: LINEAR
 - (ii) MOLECULE TYPE: PROTEIN
 - (vi) ORIGINAL SOURCE:
 - (A) ORGANISM: Homo Sapiens
 - (D) DEVELOPMENTAL STAGE: Fetal
 - (F) TISSUE TYPE: liver
 - (ix) FEATURE:
 - (A) NAME/KEY: sig_peptide
 - (B) LOCATION: -49..-1
 - (C) IDENTIFICATION METHOD: Von Heijne matrix
 - (D) OTHER INFORMATION: score 4.9 seq LLCLSGLELEPSA/SD
 - (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 243:

Met Lys Lys Leu Arg Pro Ser Gln Glu Gln Leu Asn Cys Pro Glu Pro
-45 -40 -35

Gln Leu Ala Asp Gly Arg Ala Gly Ile Arg Leu Leu Val Thr Trp Leu
-30 -25 -20

Gln Pro Ala Pro Leu Leu Cys Leu Ser Gly Leu Glu Leu Glu Pro Ser
-15 -10 -5

Ala Ser Asp Phe Gly Phe Ser Ser His Thr Thr Leu Leu Cys Cys Leu
1 5 10 15

Val Glu Asn

- (2) INFORMATION FOR SEQ ID NO: 244:
 - (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 74 amino acids
 - (B) TYPE: AMINO ACID
 - (D) TOPOLOGY: LINEAR
 - (ii) MOLECULE TYPE: PROTEIN
 - (vi) ORIGINAL SOURCE:
 - (A) ORGANISM: Homo Sapiens
 - (F) TISSUE TYPE: Large intestine
 - (ix) FEATURE:
 - (A) NAME/KEY: sig_peptide
 - (B) LOCATION: -19..-1

. 204

- (C) IDENTIFICATION METHOD: Von Heijne matrix
- (D) OTHER INFORMATION: score 4.9 seq RLLFWSIFSSVTC/RK
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 244:

Met Trp Ser His Leu Asn Arg Leu Leu Phe Trp Ser Ile Phe Ser Ser -15 -10 -5

Val Thr Cys Arg Lys Ala Val Leu Asp Cys Glu Ala Met Lys Thr Asn
1 5 10

Glu Phe Pro Ser Pro Cys Leu Asp Ser Lys Thr Lys Val Val Met Lys
15 20 25

Gly Gln Asn Val Ser Met Phe Cys Ser His Lys Asn Lys Ser Leu Gln 30 35 40 45

Ile Thr Tyr Ser Leu Phe Arg Arg Lys Thr 50 55

- (2) INFORMATION FOR SEQ ID NO: 245:
 - (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 83 amino acids
 - (B) TYPE: AMINO ACID
 - (D) TOPOLOGY: LINEAR
 - (ii) MOLECULE TYPE: PROTEIN
 - (vi) ORIGINAL SOURCE:
 - (A) ORGANISM: Homo Sapiens
 - (F) TISSUE TYPE: Lung (cells)
 - (ix) FEATURE:
 - (A) NAME/KEY: sig_peptide
 - (B) LOCATION: -52..-1
 - (C) IDENTIFICATION METHOD: Von Heijne matrix
 - (D) OTHER INFORMATION: score 4.8 seq SLLLAQATSNVVC/SL
 - (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 245:

Met Leu Ala Leu Arg Asp Leu Gly Met Gly Lys Arg Glu Gly Glu Glu
-50 -45 -40

Leu Ile Gln Ala Glu Ala Arg Cys Leu Val Glu Thr Phe Gln Gly Thr
-35 -25

Glu Gly Arg Pro Phe Asp Pro Ser Leu Leu Leu Ala Gln Ala Thr Ser
-20 -15 -10 . -5

Asn Val Val Cys Ser Leu Leu Phe Gly Leu Arg Phe Ser Tyr Glu Asp $1 \hspace{1.5cm} 5 \hspace{1.5cm} 10$

Lys Glu Phe Gln Ala Val Val Arg Ala Gly Gly Thr Cys Trp Glu 15 20 25 Ser Ala Pro 30

- (2) INFORMATION FOR SEQ ID NO: 246:
 - (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 67 amino acids
 - (B) TYPE: AMINO ACID
 - (D) TOPOLOGY: LINEAR
 - (ii) MOLECULE TYPE: PROTEIN
 - (vi) ORIGINAL SOURCE:
 - (A) ORGANISM: Homo Sapiens
 - (F) TISSUE TYPE: Liver
 - (ix) FEATURE:
 - (A) NAME/KEY: sig_peptide
 - (B) LOCATION: -19..-1
 - (C) IDENTIFICATION METHOD: Von Heijne matrix
 - (D) OTHER INFORMATION: score 4.8

seq STSLCGCLRQLRC/SM

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 246:

Met Leu Ser Val Gly Ala Ser Thr Ser Leu Cys Gly Cys Leu Arg Gln
-15 -10 -5

Leu Arg Cys Ser Met Leu Asp Leu Gln Trp Ser Phe Leu Glu Asp Gly
1 5 10

Glu Pro Cys Arg Ala Arg Leu Ser Pro Leu Pro Pro Leu Ala His Leu 15 20 25

Ala Gly Ile Trp Ile Val Leu Pro Arg Ala Ser Phe Ser Val Met Asp 30 40 45

Tyr His Ala

- (2) INFORMATION FOR SEQ ID NO: 247:
 - (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 23 amino acids
 - (B) TYPE: AMINO ACID
 - (D) TOPOLOGY: LINEAR
 - (ii) MOLECULE TYPE: PROTEIN
 - (vi) ORIGINAL SOURCE:
 - (A) ORGANISM: Homo Sapiens
 - (F) TISSUE TYPE: Lung
 - (ix) FEATURE:
 - (A) NAME/KEY: sig_peptide
 - (B) LOCATION: -19..-1

206

- (C) IDENTIFICATION METHOD: Von Heijne matrix
- (D) OTHER INFORMATION: score 4.8

seq VLLSQFLYPLAYP/HP

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 247:

Met Phe Gln Gln Met Tyr Val Leu Leu Ser Gln Phe Leu Tyr Pro Leu
-15 -10 -5

Ala Tyr Pro His Pro Ile Gly
1

- (2) INFORMATION FOR SEQ ID NO: 248:
 - (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 24 amino acids
 - (B) TYPE: AMINO ACID
 - (D) TOPOLOGY: LINEAR
 - (ii) MOLECULE TYPE: PROTEIN
 - (vi) ORIGINAL SOURCE:
 - (A) ORGANISM: Homo Sapiens
 - (F) TISSUE TYPE: Lung (cells)
 - (ix) FEATURE:
 - (A) NAME/KEY: sig_peptide
 - (B) LOCATION: -16..-1
 - (C) IDENTIFICATION METHOD: Von Heijne matrix
 - (D) OTHER INFORMATION: score 4.8

seq HFCXIGFLSYTTS/LV

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 248:

Met Thr Ser His Phe Cys Xaa Ile Gly Phe Leu Ser Tyr Thr Thr Ser
-15 -5

Leu Val Tyr Trp Asn Ala Gly Arg 1 5

- (2) INFORMATION FOR SEQ ID NO: 249:
 - (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 47 amino acids
 - (B) TYPE: AMINO ACID
 - (D) TOPOLOGY: LINEAR
 - (ii) MOLECULE TYPE: PROTEIN
 - (vi) ORIGINAL SOURCE:
 - (A) ORGANISM: Homo Sapiens
 - (F) TISSUE TYPE: Thyroid
 - (ix) FEATURE:
 - (A) NAME/KEY: sig_peptide

207

- (B) LOCATION: -25..-1
- (C) IDENTIFICATION METHOD: Von Heijne matrix
- (D) OTHER INFORMATION: score 4.8

seq NVLLSGSLLRSLC/LK

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 249:

Met Ile Cys Ser Leu Thr Pro Phe Arg Ser Leu Thr Asn Val Leu Leu -25 -15 -10

Ser Gly Ser Leu Leu Arg Ser Leu Cys Leu Lys Tyr Lys Pro Leu Thr
-5 1 5

Ser Ile Phe Leu Val Ser Met Cys Pro Ile Pro Phe Pro Cys His 10 15 20

- (2) INFORMATION FOR SEQ ID NO: 250:
 - (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 34 amino acids
 - (B) TYPE: AMINO ACID
 - (D) TOPOLOGY: LINEAR
 - (ii) MOLECULE TYPE: PROTEIN
 - (vi) ORIGINAL SOURCE:
 - (A) ORGANISM: Homo Sapiens
 - (F) TISSUE TYPE: Colon
 - (ix) FEATURE:
 - (A) NAME/KEY: sig_peptide
 - (B) LOCATION: -27..-1
 - (C) IDENTIFICATION METHOD: Von Heijne matrix
 - (D) OTHER INFORMATION: score 4.7

seq PALTLTFLPPSPT/LP

- (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 250:
- Met Glu Pro Pro Gly Arg Ser Ser Ser Leu Pro Phe Ser Pro Pro Ala -25 -20 -15

Leu Thr Leu Thr Phe Leu Pro Pro Ser Pro Thr Leu Pro Leu Pro Ser -10 5

Pro Gly

- (2) INFORMATION FOR SEQ ID NO: 251:
 - (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 63 amino acids
 - (B) TYPE: AMINO ACID
 - (D) TOPOLOGY: LINEAR
 - (ii) MOLECULE TYPE: PROTEIN

208

(vi) ORIGINAL SOURCE:

(A) ORGANISM: Homo Sapiens

(F) TISSUE TYPE: Colon

(ix) FEATURE:

(A) NAME/KEY: sig_peptide

(B) LOCATION: -55..-1

- (C) IDENTIFICATION METHOD: Von Heijne matrix
- (D) OTHER INFORMATION: score 4.7

seq IGILCSLLGTVLL/WV

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 251:

Met Asp Lys Leu Lys Lys Val Leu Ser Gly Gln Asp Thr Glu Asp Arg
-55 -50 -45 -45

Ser Gly Leu Ser Glu Val Val Glu Ala Ser Ser Leu Ser Trp Ser Thr
-35
-30
-25

Arg Ile Lys Gly Phe Ile Ala Cys Phe Ala Ile Gly Ile Leu Cys Ser
-20 -15 -10

Leu Leu Gly Thr Val Leu Leu Trp Val Pro Arg Lys Gly His Gly
-5 1 5

- (2) INFORMATION FOR SEQ ID NO: 252:
 - (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 32 amino acids
 - (B) TYPE: AMINO ACID
 - (D) TOPOLOGY: LINEAR
 - (ii) MOLECULE TYPE: PROTEIN
 - (vi) ORIGINAL SOURCE:
 - (A) ORGANISM: Homo Sapiens
 - (F) TISSUE TYPE: Lung (cells)
 - (ix) FEATURE:
 - (A) NAME/KEY: sig_peptide
 - (B) LOCATION: -25..-1
 - (C) IDENTIFICATION METHOD: Von Heijne matrix
 - (D) OTHER INFORMATION: score 4.7

seq LGMVCIFSLRLQA/VF

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 252:

Met Tyr Ser Arg His Thr Val Lys Leu Lys Gln Gly Leu Gly Met Val -25 -15 -10

Cys Ile Phe Ser Leu Arg Leu Gln Ala Val Phe Thr Thr Glu Gly Arg
-5 1 5

(2) INFORMATION FOR SEQ ID NO: 253:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 55 amino acids
 - (B) TYPE: AMINO ACID
 - (D) TOPOLOGY: LINEAR
- (ii) MOLECULE TYPE: PROTEIN
- (vi) ORIGINAL SOURCE:
 - (A) ORGANISM: Homo Sapiens
 - (F) TISSUE TYPE: Lung (cells)
- (ix) FEATURE:
 - (A) NAME/KEY: sig_peptide
 - (B) LOCATION: -24..-1
 - (C) IDENTIFICATION METHOD: Von Heijne matrix
 - (D) OTHER INFORMATION: score 4.6

seq SLLLYSLPLNIIG/LN

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 253:

Met Tyr Pro Ser Leu Leu Val Asp Tyr Phe Pro Ser Leu Leu Tyr
-20 -15 -10

Ser Leu Pro Leu Asn Ile Ile Gly Leu Asn Cys Ala Tyr Pro Leu Ile
-5 1 5

Asn Asn Phe Leu Lys Asn Asn Ser Tyr Thr Cys Val Xaa Val Pro Leu . 10 20

Ala Phe Pro Ser Met Pro Ser 25 30

- (2) INFORMATION FOR SEQ ID NO: 254:
 - (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 79 amino acids
 - (B) TYPE: AMINO ACID
 - (D) TOPOLOGY: LINEAR
 - (ii) MOLECULE TYPE: PROTEIN
 - (vi) ORIGINAL SOURCE:
 - (A) ORGANISM: Homo Sapiens
 - (F) TISSUE TYPE: Lung (cells)
 - (ix) FEATURE:
 - (A) NAME/KEY: sig_peptide
 - (B) LOCATION: -77..-1
 - (C) IDENTIFICATION METHOD: Von Heijne matrix
 - (D) OTHER INFORMATION: score 4.5

seq LPTQFLFLLGVLG/IF

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 254:

Met Ala Thr Thr Val Pro Asp Gly Cys Arg Asn Gly Leu Lys Ser Lys
-75
-70
-65

210

Tyr Tyr Arg Leu Cys Asp Lys Ala Glu Ala Trp Gly Ile Val Leu Glu
-60 -55 -50

Thr Val Ala Thr Ala Gly Val Val Thr Ser Val Ala Phe Met Xaa Thr -45 -35 -30

Leu Pro Ile Leu Val Cys Lys Val Gln Asp Ser Asn Arg Arg Lys Met
-25 -20 -15

Leu Pro Thr Gln Phe Leu Phe Leu Leu Gly Val Leu Gly Ile Phe
-10 -5 1

- (2) INFORMATION FOR SEQ ID NO: 255:
 - (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 68 amino acids
 - (B) TYPE: AMINO ACID
 - (D) TOPOLOGY: LINEAR
 - (ii) MOLECULE TYPE: PROTEIN
 - (vi) ORIGINAL SOURCE:
 - (A) ORGANISM: Homo Sapiens
 - (F) TISSUE TYPE: Large intestine
 - (ix) FEATURE:
 - (A) NAME/KEY: sig_peptide
 - (B) LOCATION: -17..-1
 - (C) IDENTIFICATION METHOD: Von Heijne matrix
 - (D) OTHER INFORMATION: score 4.5 seq HLDHLFFSGVVLG/QG
 - (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 255:

Met Arg Leu Gln His Leu Asp His Leu Phe Phe Ser Gly Val Val Leu
-15 -10 -5

Gly Gln Gly Leu Asp Leu Gly Arg Val Cys Leu Arg Lys Trp Gly Tyr
1 5 10 15

Arg Arg Cys Glu Asp Ile Cys Trp Ile Lys Thr Asn Lys Asn Asn Pro 20 25 30

Gly Lys Thr Lys Thr Leu Asp Pro Lys Ala Val Phe Gln Arg Thr Lys 35 40 45

Ala Gly Leu Gly 50

- (2) INFORMATION FOR SEQ ID NO: 256:
 - (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 78 amino acids
 - (B) TYPE: AMINO ACID
 - (D) TOPOLOGY: LINEAR

211

- (ii) MOLECULE TYPE: PROTEIN
- (vi) ORIGINAL SOURCE:
 - (A) ORGANISM: Homo Sapiens
 - (F) TISSUE TYPE: Pancreas
- (ix) FEATURE:
 - (A) NAME/KEY: sig_peptide
 - (B) LOCATION: -31..-1
 - (C) IDENTIFICATION METHOD: Von Heijne matrix
 - (D) OTHER INFORMATION: score 4.5

seq VAFGLYNPSLCHA/CT

- (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 256:
- Met Pro Leu Pro Lys Pro Ser Phe Ser Asn Asn His Leu Ile Arg Leu
 -30 -25 -20
- Ile Thr Val Ala Phe Gly Leu Tyr Asn Pro Ser Leu Cys His Ala Cys -15 -10 -5 1
- Thr Arg Cys Ser Thr Ala Ser Val Ser His Gln Ile Ala His Ser Pro 5 10 15
- Lys Gln Lys Pro Ser Asn Leu Gly Ala Ile Gln Gly Leu Ala Gln Cys 20 25 30
- Leu Val Glu His Met Cys Cys Arg Ile Asn Ile Asp Thr Trp 35 40 45
- (2) INFORMATION FOR SEQ ID NO: 257:
 - (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 94 amino acids
 - (B) TYPE: AMINO ACID
 - (D) TOPOLOGY: LINEAR
 - (ii) MOLECULE TYPE: PROTEIN
 - (vi) ORIGINAL SOURCE:
 - (A) ORGANISM: Homo Sapiens
 - (F) TISSUE TYPE: Pancreas
 - (ix) FEATURE:
 - (A) NAME/KEY: sig_peptide
 - (B) LOCATION: -75..-1
 - (C) IDENTIFICATION METHOD: Von Heijne matrix
 - (D) OTHER INFORMATION: score 4.4

seq QXRLCVSPSGLRC/EP

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 257:

Met Glu Pro Ile Thr Phe Thr Ala Arg Lys His Leu Leu Pro Asn Glu
-75 -65 -65

Val Ser Val Asp Phe Gly Leu Gln Leu Val Gly Ser Leu Pro Val His
-55 -50 -45

Ser Leu Thr Thr Met Pro Met Leu Pro Trp Val Val Ala Glu Val Arg
-40 -35 -30

Arg Leu Ser Arg Gln Ser Thr Arg Lys Glu Pro Val Thr Xaa Gln Xaa -25 -20 -15

Arg Leu Cys Val Ser Pro Ser Gly Leu Arg Cys Glu Pro Glu Pro Gly -10 -5 1 5

Arg Ser Gln Gln Trp Asp Pro Leu Ile Tyr Ser Ser Ile Phe
10 15

- (2) INFORMATION FOR SEQ ID NO: 258:
 - (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 40 amino acids
 - (B) TYPE: AMINO ACID
 - (D) TOPOLOGY: LINEAR
 - (ii) MOLECULE TYPE: PROTEIN
 - (vi) ORIGINAL SOURCE:
 - (A) ORGANISM: Homo Sapiens
 - (F) TISSUE TYPE: Colon
 - (ix) FEATURE:
 - (A) NAME/KEY: sig_peptide
 - (B) LOCATION: -25..-1
 - (C) IDENTIFICATION METHOD: Von Heijne matrix
 - (D) OTHER INFORMATION: score 4.4 seq FFCWEVGVSGSSA/GP
 - (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 258:

Met Gly Cys Leu Trp Gly Leu Ala Leu Pro Leu Phe Phe Cys Trp
-25
-10
-15

Glu Val Gly Val Ser Gly Ser Ser Ala Gly Pro Ser Thr Arg Arg Ala
-5 1 5

Asp Thr Ala Met Thr Thr Asp Asp 10 15

- (2) INFORMATION FOR SEQ ID NO: 259:
 - (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 38 amino acids
 - (B) TYPE: AMINO ACID
 - (D) TOPOLOGY: LINEAR
 - (ii) MOLECULE TYPE: PROTEIN
 - (vi) ORIGINAL SOURCE:
 - (A) ORGANISM: Homo Sapiens
 - (F) TISSUE TYPE: Lung (cells)

213

(ix) FEATURE:

(A) NAME/KEY: sig_peptide

(B) LOCATION: -20..-1

- (C) IDENTIFICATION METHOD: Von Heijne matrix
- (D) OTHER INFORMATION: score 4.3 seq YLCHISLLDVTQQ/FP

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 259:

Met Lys Gln Asn Thr Asp Pro Tyr Leu Cys His Ile Ser Leu Leu Asp
-20 -15 -10 -5

Val Thr Gln Gln Phe Pro Asn Pro Leu Pro Gly Arg Thr Ile Phe Pro 1 5 10

Gly Ser Ser Thr Pro Arg 15

- (2) INFORMATION FOR SEQ ID NO: 260:
 - (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 51 amino acids
 - (B) TYPE: AMINO ACID
 - (D) TOPOLOGY: LINEAR
 - (ii) MOLECULE TYPE: PROTEIN
 - (vi) ORIGINAL SOURCE:
 - (A) ORGANISM: Homo Sapiens
 - (D) DEVELOPMENTAL STAGE: Fetal
 - (F) TISSUE TYPE: liver
 - (ix) FEATURE:
 - (A) NAME/KEY: sig_peptide
 - (B) LOCATION: -35..-1
 - (C) IDENTIFICATION METHOD: Von Heijne matrix
 - (D) OTHER INFORMATION: score 4.3

seq LISLLSSPNTPSA/SI

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 260:

Met Val Thr Tyr Phe Asn Phe Thr Phe Lys Pro Phe Cys Ile Leu Ala
-35 -25 -20

Ser Ile Ile Val Pro Thr Leu Ile Ser Leu Leu Ser Ser Pro Asn Thr
-15 -10 -5

Pro Ser Ala Ser Ile Tyr Tyr Ser Pro Lys Cys Leu Cys Pro Leu Ala 1 5 10

Thr Pro Arg

(2) INFORMATION FOR SEQ ID NO: 261:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 28 amino acids
- (B) TYPE: AMINO ACID
- (D) TOPOLOGY: LINEAR
- (ii) MOLECULE TYPE: PROTEIN
- (vi) ORIGINAL SOURCE:
 - (A) ORGANISM: Homo Sapiens
 - (F) TISSUE TYPE: Colon
- (ix) FEATURE:
 - (A) NAME/KEY: sig_peptide
 - (B) LOCATION: -23..-1
 - (C) IDENTIFICATION METHOD: Von Heijne matrix
 - (D) OTHER INFORMATION: score 4.3

seq QXILLGTTSVVTA/AL

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 261:

Met Glu Ser Gly Gly Arg Pro Ser Leu Cys Gln Xaa Ile Leu Leu Gly
-20 -15 -10

Thr Thr Ser Val Val Thr Ala Ala Leu T∳r Ser Val
-5 1 5

- (2) INFORMATION FOR SEQ ID NO: 262:
 - (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 91 amino acids
 - (B) TYPE: AMINO ACID
 - (D) TOPOLOGY: LINEAR
 - (ii) MOLECULE TYPE: PROTEIN
 - (vi) ORIGINAL SOURCE:
 - (A) ORGANISM: Homo Sapiens
 - (F) TISSUE TYPE: Pancreas
 - (ix) FEATURE:
 - (A) NAME/KEY: sig_peptide
 - (B) LOCATION: -43..-1
 - (C) IDENTIFICATION METHOD: Von Heijne matrix
 - (D) OTHER INFORMATION: score 4.2

seq HMMAAAVADGTRA/AT

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 262:

Met Glu Ala Gln Gln Ala Gln Lys Ser Ala Glu Gln Pro Glu Gln Lys
-40 -35 -30

Ala Ala Thr Glu Val Ser Xaa Glu Leu Ser Glu Ser Gln Val His Met
-25 -20 -15

Met Ala Ala Val Ala Asp Gly Thr Arg Ala Ala Thr Ile Ile Glu
-10 -5 1

215

Glu Arg Ser Pro Ser Trp Ile Ser Ala Ser Val Thr Glu Pro Leu Glu

Gln Val Glu Ala Glu Ala Ala Leu Leu Thr Glu Glu Val Leu Glu Arg

Glu Val Ile Ala Glu Glu Glu Pro Pro Thr Met 40

- (2) INFORMATION FOR SEQ ID NO: 263:
 - (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 21 amino acids
 - (B) TYPE: AMINO ACID
 - (D) TOPOLOGY: LINEAR
 - (ii) MOLECULE TYPE: PROTEIN
 - (vi) ORIGINAL SOURCE:
 - (A) ORGANISM: Homo Sapiens
 - (F) TISSUE TYPE: Lung (cells)
 - (ix) FEATURE:
 - (A) NAME/KEY: sig_peptide
 - (B) LOCATION: -17..-1
 - (C) IDENTIFICATION METHOD: Von Heijne matrix
 - (D) OTHER INFORMATION: score 4.2

seq SVIWFGSVXPCIS/XV

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 263:

Met Pro Leu Asn Ser Val Ile Trp Phe Gly Ser Val Xaa Pro Cys Ile -15 -10

Ser Xaa Val Glu Leu 1

- (2) INFORMATION FOR SEQ ID NO: 264:
 - (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 37 amino acids
 - (B) TYPE: AMINO ACID
 - (D) TOPOLOGY: LINEAR
 - (ii) MOLECULE TYPE: PROTEIN
 - (vi) ORIGINAL SOURCE:
 - (A) ORGANISM: Homo Sapiens
 - (F) TISSUE TYPE: Lung
 - (ix) FEATURE:
 - (A) NAME/KEY: sig_peptide
 - (B) LOCATION: -30..-1
 - (C) IDENTIFICATION METHOD: Von Heijne matrix
 - (D) OTHER INFORMATION: score 4.2

seq FLDFANLADLTLA/ES

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 264:

Met Leu Gln Gln Leu Asp Ser Ile Ser Leu Arg Arg Glu Thr Ala
-30 -25 -20 -15

Asn Phe Leu Asp Phe Ala Asn Leu Ala Asp Leu Thr Leu Ala Glu Ser
-10 -5 1

Glu Val Phe Arg Leu 5

- (2) INFORMATION FOR SEQ ID NO: 265:
 - (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 56 amino acids
 - (B) TYPE: AMINO ACID
 - (D) TOPOLOGY: LINEAR
 - (ii) MOLECULE TYPE: PROTEIN
 - (vi) ORIGINAL SOURCE:
 - (A) ORGANISM: Homo Sapiens
 - (F) TISSUE TYPE: Lung (cells)
 - (ix) FEATURE:
 - (A) NAME/KEY: sig_peptide
 - (B) LOCATION: -48..-1
 - (C) IDENTIFICATION METHOD: Von Heijne matrix
 - (D) OTHER INFORMATION: score 4.1

seq FTTLSNLSLPSQT/KD

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 265:

Met Cys Tyr Leu Ala Glu Leu Ser Leu Thr Thr Phe Xaa Xaa Gly Tyr
-45 -40 -35

Ile Val Thr Ser Arg Ala Thr Thr Thr Thr Thr Leu Ala Ile Gln Pro
-30 -25 -20

Gly Leu Pro Phe Thr Thr Leu Ser Asn Leu Ser Leu Pro Ser Gln Thr
-15 -10 -5

Lys Asp Glu Leu His Pro Pro Trp

- (2) INFORMATION FOR SEQ ID NO: 266:
 - (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 24 amino acids
 - (B) TYPE: AMINO ACID
 - (D) TOPOLOGY: LINEAR
 - (ii) MOLECULE TYPE: PROTEIN

(vi) ORIGINAL SOURCE:

(A) ORGANISM: Homo Sapiens

(F) TISSUE TYPE: Pancreas

(ix) FEATURE:

- (A) NAME/KEY: sig_peptide
- (B) LOCATION: -17..-1
- (C) IDENTIFICATION METHOD: Von Heijne matrix
- (D) OTHER INFORMATION: score 4.1

seq LSSLILLPIWINM/AQ

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 266:

Met Ser Ile Ser Leu Ser Ser Leu Ile Leu Leu Pro Ile Trp Ile Asn
-15 -10 -5

Met Ala Gln Ile Gln Arg Gly Gly

- (2) INFORMATION FOR SEQ ID NO: 267:
 - (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 125 amino acids
 - (B) TYPE: AMINO ACID
 - (D) TOPOLOGY: LINEAR
 - (ii) MOLECULE TYPE: PROTEIN
 - (vi) ORIGINAL SOURCE:
 - (A) ORGANISM: Homo Sapiens
 - (F) TISSUE TYPE: Lung
 - (ix) FEATURE:
 - (A) NAME/KEY: sig_peptide
 - (B) LOCATION: -86..-1
 - (C) IDENTIFICATION METHOD: Von Heijne matrix
 - (D) OTHER INFORMATION: score 4

seq LLFALSWKSDAPA/TS

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 267:

Met Asp Arg Asp Leu Leu Arg Gln Ser Leu Asn Cys His Gly Ser Ser
-85 -80 -75

Leu Leu Ser Leu Leu Arg Ser Glu Gln Gln Asp Asn Pro His Phe Arg
-70 -65 -60 -55

Ser Leu Leu Gly Ser Ala Ala Glu Pro Ala Arg Gly Pro Pro Gln
-50 -45 -40

His Pro Leu Gln Gly Arg Lys Glu Lys Arg Val Asp Asn Ile Glu Ile
-35 -30 -25

Gln Lys Phe Ile Ser Lys Lys Ala Asp Leu Leu Phe Ala Leu Ser Trp
-20 -15 -10

218

Lys Ser Asp Ala Pro Ala Thr Ser Glu Ile Asn Glu Asp Ser Glu Asp
-5 1 5 10

His Tyr Ala Ile Met Pro Pro Leu Glu Gln Phe Met Glu Ile Pro Ser 15 20 25

Met Asp Arg Arg Glu Leu Phe Phe Arg Asp Ile Glu Arg 30 35

- (2) INFORMATION FOR SEQ ID NO: 268:
 - (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 25 amino acids
 - (B) TYPE: AMINO ACID
 - (D) TOPOLOGY: LINEAR
 - (ii) MOLECULE TYPE: PROTEIN
 - (vi) ORIGINAL SOURCE:
 - (A) ORGANISM: Homo Sapiens
 - (F) TISSUE TYPE: Large intestine
 - (ix) FEATURE:
 - (A) NAME/KEY: sig_peptide
 - (B) LOCATION: -17..-1
 - (C) IDENTIFICATION METHOD: Von Heijne matrix
 - (D) OTHER INFORMATION: score 4 seq TLVTXXNASCSFA/SV
 - (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 268:

Met Val Leu Ala Thr Leu Val Thr Xaa Xaa Asn Ala Ser Cys Ser Phe
-15 -10 -5

Ala Ser Val His Leu Ala Gln Gly Gly
1 5

- (2) INFORMATION FOR SEQ ID NO: 269:
 - (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 39 amino acids
 - (B) TYPE: AMINO ACID
 - (D) TOPOLOGY: LINEAR
 - (ii) MOLECULE TYPE: PROTEIN
 - (vi) ORIGINAL SOURCE:
 - (A) ORGANISM: Homo Sapiens
 - (F) TISSUE TYPE: Lung (cells)
 - (ix) FEATURE:
 - (A) NAME/KEY: sig_peptide
 - (B) LOCATION: -20..-1
 - (C) IDENTIFICATION METHOD: Von Heijne matrix
 - (D) OTHER INFORMATION: score 4

seg IILKVLLNQTCQT/VO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 269:

Met Met Ile Trp Lys Arg Leu Ile Ile Leu Lys Val Leu Leu Asn Gln
-20 -15 -10 -5

Thr Cys Gln Thr Val Gln Thr Val Thr Pro Thr Ser Trp Val Phe Ser $1 \hspace{1cm} 5 \hspace{1cm} 10$

Asn Gln Ala Gly Met Thr Arg 15

- (2) INFORMATION FOR SEQ ID NO: 270:
 - (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 100 amino acids
 - (B) TYPE: AMINO ACID
 - (D) TOPOLOGY: LINEAR
 - (ii) MOLECULE TYPE: PROTEIN
 - (vi) ORIGINAL SOURCE:
 - (A) ORGANISM: Homo Sapiens
 - (D) DEVELOPMENTAL STAGE: Fetal
 - (F) TISSUE TYPE: liver
 - (ix) FEATURE:
 - (A) NAME/KEY: sig_peptide
 - (B) LOCATION: -61..-1
 - (C) IDENTIFICATION METHOD: Von Heijne matrix
 - (D) OTHER INFORMATION: score 4

seq AALVKCLPVLCLA/GF

- (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 270:
- Met Asp Ala Gly Lys Ala Gly Gln Thr Leu Lys Thr His Cys Ser Ala
 -60 -55 -50

Gln Arg Pro Asp Val Cys Arg Trp Leu Ser Pro Phe Ile Leu Ser Cys
-45 -35 -30

Cys Val Tyr Phe Cys Leu Trp Ile Pro Glu Asp Gln Leu Ser Trp Phe
-25 -20 -15

Ala Ala Leu Val Lys Cys Leu Pro Val Leu Cys Leu Ala Gly Phe Leu
-10 -5 1

Trp Val Met Ser Pro Ser Gly Gly Tyr Thr Gln Leu Leu Gln Gly Ala
5 10 15

Leu Val Cys Ser Ala Val Gly Asp Ala Cys Leu Ile Trp Pro Ala Ala 20 25 30 35

Phe Val Pro Gly

- (2) INFORMATION FOR SEQ ID NO: 271:
 - (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 70 amino acids
 - (B) TYPE: AMINO ACID
 - (D) TOPOLOGY: LINEAR
 - (ii) MOLECULE TYPE: PROTEIN
 - (vi) ORIGINAL SOURCE:
 - (A) ORGANISM: Homo Sapiens
 - (F) TISSUE TYPE: Lung (cells)
 - (ix) FEATURE:
 - (A) NAME/KEY: sig_peptide
 - (B) LOCATION: -28..-1
 - (C) IDENTIFICATION METHOD: Von Heijne matrix
 - (D) OTHER INFORMATION: score 4

seq PLLGVLFFQGVYI/VF

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 271:

Met Gln Gln Arg Gly Ala Ala Gly Ser Arg Gly Cys Ala Leu Phe Pro
-25 -20 -15

Leu Leu Gly Val Leu Phe Phe Gln Gly Val Tyr Ile Val Phe Ser Leu
-10 -5 1

Glu Ile Arg Ala Asp Ala His Val Arg Gly Tyr Val Gly Glu Lys Ile 5 10 15 20

Lys Leu Lys Cys Thr Phe Lys Ser Thr Ser Asp Val Thr Asp Lys Leu 25 30 . 35

Thr Ile Asp Trp Thr Gln 40

- (2) INFORMATION FOR SEQ ID NO: 272:
 - (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 51 amino acids
 - (B) TYPE: AMINO ACID
 - (D) TOPOLOGY: LINEAR
 - (ii) MOLECULE TYPE: PROTEIN
 - (vi) ORIGINAL SOURCE:
 - (A) ORGANISM: Homo Sapiens
 - (F) TISSUE TYPE: Colon
 - (ix) FEATURE:
 - (A) NAME/KEY: sig_peptide
 - (B) LOCATION: -32..-1
 - (C) IDENTIFICATION METHOD: Von Heijne matrix
 - (D) OTHER INFORMATION: score 4 seq LIYWYVLLILSFP/FI
 - (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 272:

PCT/IB98/01233

221

Met Leu Gly Thr His Ile Tyr Val Ser Leu Trp Ile Ile Leu Phe Ser
-30 -25 -20

Ser Pro His Leu Ile Tyr Trp Tyr Val Leu Leu Ile Leu Ser Phe Pro
-15 -5

Phe Ile Ile Lys Phe Ser Met Asn Thr Leu Ser Arg Pro Pro Pro Asp 1 5 10 15

Thr Pro Gln

- (2) INFORMATION FOR SEQ ID NO: 273:
 - (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 21 amino acids
 - (B) TYPE: AMINO ACID
 - (D) TOPOLOGY: LINEAR
 - (ii) MOLECULE TYPE: PROTEIN
 - (vi) ORIGINAL SOURCE:
 - (A) ORGANISM: Homo Sapiens
 - (F) TISSUE TYPE: Pancreas
 - (ix) FEATURE:
 - (A) NAME/KEY: sig_peptide
 - (B) LOCATION: -19..-1
 - (C) IDENTIFICATION METHOD: Von Heijne matrix
 - (D) OTHER INFORMATION: score 3.9

seq FLNLHGFLGHLLS/GE

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 273:

Met Ser Ile Tyr Asn Leu Phe Leu Asn Leu His Gly Phe Leu Gly His -15 -10 -5

Leu Leu Ser Gly Glu

- (2) INFORMATION FOR SEQ ID NO: 274:
 - (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 29 amino acids
 - (B) TYPE: AMINO ACID
 - (D) TOPOLOGY: LINEAR
 - (ii) MOLECULE TYPE: PROTEIN
 - (vi) ORIGINAL SOURCE:
 - (A) ORGANISM: Homo Sapiens
 - (F) TISSUE TYPE: Lung (cells)
 - (ix) FEATURE:
 - (A) NAME/KEY: sig_peptide

- (B) LOCATION: -25..-1
- (C) IDENTIFICATION METHOD: Von Heijne matrix
- (D) OTHER INFORMATION: score 3.9

seq PACVCMCTXSCYS/CQ

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 274:

Met Cys Met Gln Val Asp Leu Ala Phe Ser Phe Pro Pro Ala Cys Val -25 -15 -10

Cys Met Cys Thr Xaa Ser Cys Tyr Ser Cys Glu -5

- (2) INFORMATION FOR SEQ ID NO: 275:
 - (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 70 amino acids
 - (B) TYPE: AMINO ACID
 - (D) TOPOLOGY: LINEAR
 - (ii) MOLECULE TYPE: PROTEIN
 - (vi) ORIGINAL SOURCE:
 - (A) ORGANISM: Homo Sapiens
 - (F) TISSUE TYPE: Pancreas
 - (ix) FEATURE:
 - (A) NAME/KEY: sig_peptide
 - (B) LOCATION: -32..-1
 - (C) IDENTIFICATION METHOD: Von Heijne matrix
 - (D) OTHER INFORMATION: score 3.8

seq IRTATLVISLARG/WQ

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 275:

Met Ala Pro Gly Glu Lys Glu Ser Gly Glu Gly Pro Ala Lys Ser Ala
-30 -25 -20

Leu Arg Lys Ile Arg Thr Ala Thr Leu Val Ile Ser Leu Ala Arg Gly
-15 -5

Trp Gln Gln Trp Ala Asn Glu Asn Ser Ile Arg Gln Ala Gln Glu Pro 1 5 10 15

Thr Gly Trp Leu Pro Gly Gly Thr Gln Asp Ser Pro Gln Ala Pro Lys
20 25 30

Pro Ile Thr Pro Arg Gly 35

- (2) INFORMATION FOR SEQ ID NO: 276:
 - (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 52 amino acids
 - (B) TYPE: AMINO ACID

223

(ii) MOLECULE TYPE: PROTEIN

(D) TOPOLOGY: LINEAR

- (vi) ORIGINAL SOURCE:
 - (A) ORGANISM: Homo Sapiens
 - (D) DEVELOPMENTAL STAGE: Fetal
 - (F) TISSUE TYPE: liver
- (ix) FEATURE:
 - (A) NAME/KEY: sig peptide
 - (B) LOCATION: -42..-1
 - (C) IDENTIFICATION METHOD: Von Heijne matrix
 - (D) OTHER INFORMATION: score 3.7

seq PLLHLFYQHLCFP/VP

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 276:

Met Glu Pro Lys Arg Gly Arg Met Trp Xaa Phe Glu Ile Glu Asp Ser -40 -35 -30

Cys Ile Tyr Gln Asp Ile Pro Ser Phe Val Leu Leu Tyr Pro Leu Leu -25 -15

His Leu Phe Tyr Gln His Leu Cys Phe Pro Val Pro Cys Thr Arg Asn -10 -5 5

Pro Gly Pro Gly

- (2) INFORMATION FOR SEQ ID NO: 277:
 - (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 27 amino acids
 - (B) TYPE: AMINO ACID
 - (D) TOPOLOGY: LINEAR
 - (ii) MOLECULE TYPE: PROTEIN
 - (vi) ORIGINAL SOURCE:
 - (A) ORGANISM: Homo Sapiens
 - (F) TISSUE TYPE: Lung (cells)
 - (ix) FEATURE:
 - (A) NAME/KEY: sig_peptide
 - (B) LOCATION: -17..-1
 - (C) IDENTIFICATION METHOD: Von Heijne matrix
 - (D) OTHER INFORMATION: score 3.6

seq SVLQRCLFSFVTS/VF

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 277:

Met Glu Phe Cys Ser Val Leu Gln Arg Cys Leu Phe Ser Phe Val Thr -15 -10 -5

Ser Val Phe His Met Leu Phe Pro Leu Pro Gly $1 \hspace{1cm} 5 \hspace{1cm} 10$

224

(2) INFORMATION FOR SEQ ID NO: 278:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 23 amino acids
- (B) TYPE: AMINO ACID
- (D) TOPOLOGY: LINEAR
- (ii) MOLECULE TYPE: PROTEIN
- (vi) ORIGINAL SOURCE:
 - (A) ORGANISM: Homo Sapiens
 - (D) DEVELOPMENTAL STAGE: Fetal
 - (F) TISSUE TYPE: liver
- (ix) FEATURE:
 - (A) NAME/KEY: sig_peptide
 - (B) LOCATION: -18..-1
 - (C) IDENTIFICATION METHOD: Von Heijne matrix
 - (D) OTHER INFORMATION: score 3.6

seq IYVLLFFLLMKFS/FD

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 278:

Met Ala Glu Ser Gln Ile Tyr Val Leu Leu Phe Phe Leu Leu Met Lys

Phe Ser Phe Asp Thr Arg Gly

- (2) INFORMATION FOR SEQ ID NO: 279:
 - (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 27 amino acids
 - (B) TYPE: AMINO ACID
 - (D) TOPOLOGY: LINEAR
 - (ii) MOLECULE TYPE: PROTEIN
 - (vi) ORIGINAL SOURCE:
 - (A) ORGANISM: Homo Sapiens
 - (F) TISSUE TYPE: Liver
 - (ix) FEATURE:
 - (A) NAME/KEY: sig_peptide
 - (B) LOCATION: -18..-1
 - (C) IDENTIFICATION METHOD: Von Heijne matrix
 - (D) OTHER INFORMATION: score 3.6

seq ACSLSSGPLQINA/LP

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 279:

Met Gln Thr Asn Asn Ala Cys Ser Leu Ser Ser Gly Pro Leu Gln Ile -10

225

Asn Ala Leu Pro Asp Leu Pro Cys His Pro Gly

- (2) INFORMATION FOR SEQ ID NO: 280:
 - (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 29 amino acids
 - (B) TYPE: AMINO ACID
 - (D) TOPOLOGY: LINEAR
 - (ii) MOLECULE TYPE: PROTEIN
 - (vi) ORIGINAL SOURCE:
 - (A) ORGANISM: Homo Sapiens
 - (F) TISSUE TYPE: Large intestine
 - (ix) FEATURE:
 - (A) NAME/KEY: sig_peptide
 - (B) LOCATION: -26..-1
 - (C) IDENTIFICATION METHOD: Von Heijne matrix
 - (D) OTHER INFORMATION: score 3.6 seq KVLMGLLCNQTAA/KR
 - (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 280:

Met Gly Gln Asn Asn Ala Ser Phe His Cys Pro Cys Leu Lys Val Leu
-25 -20 -15

Met Gly Leu Cys Asn Gln Thr Ala Ala Lys Arg Pro -10 -5 1

- (2) INFORMATION FOR SEQ ID NO: 281:
 - (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 32 amino acids
 - (B) TYPE: AMINO ACID
 - (D) TOPOLOGY: LINEAR
 - (ii) MOLECULE TYPE: PROTEIN
 - (vi) ORIGINAL SOURCE:
 - (A) ORGANISM: Homo Sapiens
 - (F) TISSUE TYPE: Liver
 - (ix) FEATURE:
 - (A) NAME/KEY: sig_peptide
 - (B) LOCATION: -14..-1
 - (C) IDENTIFICATION METHOD: Von Heijne matrix
 - (D) OTHER INFORMATION: score 3.6

seq LPLLSVMWSPIAP/LT

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 281:

Met Leu Pro Leu Leu Ser Val Met Trp Ser Pro Ile Ala Pro Leu Thr
-10 -5 1

Val Gly Ser Lys Asp Pro Cys His Phe Ile Pro Val His Asp Glu Met
5 10 15

- (2) INFORMATION FOR SEQ ID NO: 282:
 - (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 31 amino acids
 - (B) TYPE: AMINO ACID
 - (D) TOPOLOGY: LINEAR
 - (ii) MOLECULE TYPE: PROTEIN
 - (vi) ORIGINAL SOURCE:
 - (A) ORGANISM: Homo Sapiens
 - (F) TISSUE TYPE: Thyroid
 - (ix) FEATURE:
 - (A) NAME/KEY: sig_peptide
 - (B) LOCATION: -23..-1
 - (C) IDENTIFICATION METHOD: Von Heijne matrix
 - (D) OTHER INFORMATION: score 3.6

seq WITCPPTFHGCRA/LF

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 282:

Met Trp Leu Asn Cys Gly Gly Leu Gln Arg Trp Ile Thr Cys Pro Pro
-20 -15 -10

Thr Phe His Gly Cys Arg Ala Leu Phe Pro Val Leu Asp Ala Gly
-5 1 5

- (2) INFORMATION FOR SEQ ID NO: 283:
 - (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 40 amino acids
 - (B) TYPE: AMINO ACID
 - (D) TOPOLOGY: LINEAR
 - (ii) MOLECULE TYPE: PROTEIN
 - (vi) ORIGINAL SOURCE:
 - (A) ORGANISM: Homo Sapiens
 - (F) TISSUE TYPE: Lung (cells)
 - (ix) FEATURE:
 - (A) NAME/KEY: sig_peptide
 - (B) LOCATION: -32..-1
 - (C) IDENTIFICATION METHOD: Von Heijne matrix
 - (D) OTHER INFORMATION: score 3.6

seq LGVLTFILQRTTC/LN

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 283:

Met Trp Gln Gly Cys Asn Cys Ser Gln Leu Ser Glu Thr Ala Val Asp

227

-30 -25 -20

Gln Glu Gln Leu Gly Val Leu Thr Phe Ile Leu Gln Arg Thr Thr Cys
-15 -10 -5

Leu Asn Val Ser Ala Gly Lys Arg 1 5

- (2) INFORMATION FOR SEQ ID NO: 284:
 - (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 42 amino acids
 - (B) TYPE: AMINO ACID
 - (D) TOPOLOGY: LINEAR
 - (ii) MOLECULE TYPE: PROTEIN
 - (vi) ORIGINAL SOURCE:
 - (A) ORGANISM: Homo Sapiens
 - (F) TISSUE TYPE: Lung (cells)
 - (ix) FEATURE:
 - (A) NAME/KEY: sig_peptide
 - (B) LOCATION: -37..-1
 - (C) IDENTIFICATION METHOD: Von Heijne matrix
 - (D) OTHER INFORMATION: score 3.6

seq LVTLLASKSPSCP/LH

- (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 284:
- Met Cys Leu Pro His Pro Gln Val Val Ser Ser Asn Phe His Ile Leu
 -35 -30 -25
- Ile Phe Leu Leu Pro Thr Lys Met Leu Val Thr Leu Leu Ala Ser Lys
 -20 -15 -10
- Ser Pro Ser Cys Pro Leu His Pro Leu Arg
- (2) INFORMATION FOR SEQ ID NO: 285:
 - (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 111 amino acids
 - (B) TYPE: AMINO ACID
 - (D) TOPOLOGY: LINEAR
 - (ii) MOLECULE TYPE: PROTEIN
 - (vi) ORIGINAL SOURCE:
 - (A) ORGANISM: Homo Sapiens
 - (F) TISSUE TYPE: Pancreas
 - (ix) FEATURE:
 - (A) NAME/KEY: sig_peptide
 - (B) LOCATION: -74..-1
 - (C) IDENTIFICATION METHOD: Von Heijne matrix

228

(D) OTHER INFORMATION: score 3.5 seq ECLNLLLSSGADL/RR

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 285:

Met His Leu Ala Val Leu Phe Xaa Phe Ser Asp Cys Cys Arg Lys Xaa
-70 -65 -60

Leu Ser Ser Gly Gln Leu Tyr Ser Ile Val Ser Ser Leu Ser Asn Glu
-55 -50 -45

His Val Leu Ser Ala Gly Phe Asp Ile Asn Thr Pro Asp Asn Leu Gly
-40 -35 -30

Arg Thr Cys Leu His Ala Ala Ala Ser Gly Gly Asn Val Glu Cys Leu
-25 -20 -15

Asn Leu Leu Ser Ser Gly Ala Asp Leu Arg Arg Asp Lys Phe -10 -5 1 5

Gly Arg Thr Pro Leu Xaa Tyr Ala Ala Ala Asn Gly Ser Xaa Gln Cys 10 15 20

Ala Val Thr Leu Val Thr Ala Gly Ala Gly Val Asn Glu Gly Xaa 25 30 35

(2) INFORMATION FOR SEQ ID NO: 286:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 33 amino acids
 - (B) TYPE: AMINO ACID
 - (D) TOPOLOGY: LINEAR
- (ii) MOLECULE TYPE: PROTEIN
- (vi) ORIGINAL SOURCE:
 - (A) ORGANISM: Homo Sapiens
 - (F) TISSUE TYPE: Thyroid
- (ix) FEATURE:
 - (A) NAME/KEY: sig_peptide
 - (B) LOCATION: -24..-1
 - (C) IDENTIFICATION METHOD: Von Heijne matrix
 - (D) OTHER INFORMATION: score 3.5

seq LHDCFLSVFQVLS/SI

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 286:

Met Ser Phe Gln Trp Cys Gly Trp Gln Trp Gly Leu His Asp Cys Phe
-20 -15 -10

Leu Ser Val Phe Gln Val Leu Ser Ser Ile Gly Leu Val Ser Phe Leu -5 1 5

P'ne

- (2) INFORMATION FOR SEQ ID NO: 287:
 - (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 57 amino acids
 - (B) TYPE: AMINO ACID
 - (D) TOPOLOGY: LINEAR
 - (ii) MOLECULE TYPE: PROTEIN
 - (vi) ORIGINAL SOURCE:
 - (A) ORGANISM: Homo Sapiens
 - (F) TISSUE TYPE: Thyroid
 - (ix) FEATURE:
 - (A) NAME/KEY: sig_peptide
 - (B) LOCATION: -20..-1
 - (C) IDENTIFICATION METHOD: Von Heijne matrix
 - (D) OTHER INFORMATION: score 3.5

seq KFCLICLLTF1FH/HC

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 287:

Met Lys Val His Met His Thr Lys Phe Cys Leu Ile Cys Leu Leu Thr
-20 -15 -10 -5

Phe Ile Phe His His Cys Asn His Cys His Glu Glu His Asp His Gly

1 5 10

Pro Glu Ala Leu His Arg Xaa His Arg Gly Met Thr Glu Leu Glu Pro 15 20 25

Ser Lys Phe Ser Lys Gln Ala Arg Gly 30 35

- (2) INFORMATION FOR SEQ ID NO: 288:
 - (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 77 amino acids
 - (B) TYPE: AMINO ACID
 - (D) TOPOLOGY: LINEAR
 - (ii) MOLECULE TYPE: PROTEIN
 - (vi) ORIGINAL SOURCE:
 - (A) ORGANISM: Homo Sapiens
 - (F) TISSUE TYPE: Lung
 - (ix) FEATURE:
 - (A) NAME/KEY: sig_peptide
 - (B) LOCATION: -25..-1
 - (C) IDENTIFICATION METHOD: Von Heijne matrix
 - (D) OTHER INFORMATION: score 3.5

seq IFLGKSLFSLLEA/MI

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 288:

230

Met Ser Phe Asn Leu Gln Ser Ser Lys Lys Leu Phe Ile Phe Leu Gly
-25 -10 -15

Lys Ser Leu Phe Ser Leu Leu Glu Ala Met Ile Phe Ala Leu Leu Pro

Lys Pro Arg Lys Asn Val Ala Gly Glu Ile Val Leu Ile Thr Gly Ala 10 15 20

Gly Ser Gly Leu Gly Arg Leu Leu Ala Leu Gln Phe Ala Arg Leu Gly 25 30 35

Ser Val Leu Val Leu Trp Asp Ile Asn Lys Glu Gly Arg

(2) INFORMATION FOR SEQ ID NO: 289:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 126 amino acids
 - (B) TYPE: AMINO ACID
 - (D) TOPOLOGY: LINEAR
- (ii) MOLECULE TYPE: PROTEIN
- (vi) ORIGINAL SOURCE:
 - (A) ORGANISM: Homo Sapiens
 - (F) TISSUE TYPE: Large intestine
- (ix) FEATURE:
 - (A) NAME/KEY: sig_peptide
 - (B) LOCATION: -26..-1
 - (C) IDENTIFICATION METHOD: Von Heijne matrix
 - (D) OTHER INFORMATION: score 7.6 seq FLLLVAAPRWVVS/EM
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 289:

Met Asp Leu Met Cys Arg Lys Val Lys His Leu Leu Phe Phe Leu Leu
-25 -20 -15

Leu Val Ala Ala Pro Arg Trp Val Val Ser Glu Met Gln Ile Glu Glu -10 -5 1 5

Ser Gly Pro Gly Leu Val Lys Pro Ser Glu Thr Leu Thr Leu Thr Cys
10 15 20

Asn Val Phe Gly Gly Ala Ile Asn Thr Asn Ala Tyr Tyr Trp Ala Trp
25 30 35

Val Arg Gln Ser Pro Gly Lys Gly Leu Glu Trp Ile Gly Ser Val Tyr 40 45 50

Tyr Ser Gly Ser Thr Tyr Tyr Asn Pro Ser Leu Lys Ser Arg Val Thr 55 60 . 65 70

Met Ser Met Ala Thr Ser Arg Asn Gln Phe Ser Leu Gln Met Ser Ser 75 30 25

231

Val Met Ala Thr Asp Thr Ala Val Tyr Tyr Cys Ala Arg Gln 90 95 100

- (2) INFORMATION FOR SEQ ID NO: 290:
 - (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 63 amino acids
 - (B) TYPE: AMINO ACID
 - (D) TOPOLOGY: LINEAR
 - (ii) MOLECULE TYPE: PROTEIN
 - (vi) ORIGINAL SOURCE:
 - (A) ORGANISM: Homo Sapiens
 - (F) TISSUE TYPE: Lung (cells)
 - (ix) FEATURE:
 - (A) NAME/KEY: sig_peptide
 - (B) LOCATION: -52..-1
 - (C) IDENTIFICATION METHOD: Von Heijne matrix
 - (D) OTHER INFORMATION: score 6.6

seq FFAVLFFLWRSFX/SV

- (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 290:
- Met Glu Leu Lys Ser Pro Glu Glu Glu Val Val Ala Ala Leu Pro Glu
 -50 -45 -40
- Gly Met Arg Pro Asp Ser Asn Leu Tyr Gly Phe Pro Trp Glu Leu Val -35 -30 -25
- Ile Cys Ala Ala Val Val Gly Phe Phe Ala Val Leu Phe Phe Leu Trp
 -20 -15 -10 -5
- Arg Ser Phe Xaa Ser Val Arg Ser Arg Leu Tyr Val Gly Arg Gly
 1 5 10
- (2) INFORMATION FOR SEQ ID NO: 291:
 - (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 77 amino acids
 - (B) TYPE: AMINO ACID
 - (D) TOPOLOGY: LINEAR
 - (ii) MOLECULE TYPE: PROTEIN
 - (vi) ORIGINAL SOURCE:
 - (A) ORGANISM: Homo Sapiens
 - (F) TISSUE TYPE: Thyroid
 - (ix) FEATURE:
 - (A) NAME/KEY: sig_peptide
 - (B) LOCATION: -39..-1
 - (C) IDENTIFICATION METHOD: Von Heijne matrix
 - (D) OTHER INFORMATION: score 6.6

seq LALVLAWLSTYVA/DS

232

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 291:

Met Glu Leu Ser Asp Val Thr Leu Ile Glu Gly Val Gly Asn Glu Val
-35 -30 -25

Met Val Val Ala Gly Val Val Leu Ile Leu Ala Leu Val Leu Ala -20 -15 -10

Trp Leu Ser Thr Tyr Val Ala Asp Ser Gly Ser Asn Gln Leu Leu Gly
-5 1 5

Ala Ile Val Ser Ala Gly Asp Thr Ser Val Leu Xaa Leu Gly His Val 10 20 25

Asp His Leu Val Ala Gly Gln Gly Asn Pro Glu Arg Arg

(2) INFORMATION FOR SEQ ID NO: 292:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 67 amino acids
 - (B) TYPE: AMINO ACID
 - (D) TOPOLOGY: LINEAR
- (ii) MOLECULE TYPE: PROTEIN
- (vi) ORIGINAL SOURCE:
 - (A) ORGANISM: Homo Sapiens
 - (F) TISSUE TYPE: Pancreas
- (ix) FEATURE:
 - (A) NAME/KEY: sig_peptide
 - (B) LOCATION: -41..-1
 - (C) IDENTIFICATION METHOD: Von Heijne matrix
 - (D) OTHER INFORMATION: score 5.4

seq RLLYIGFLGYCSG/LI

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 292:

Met Ile Ala Arg Arg Asn Pro Glu Pro Leu Arg Phe Leu Pro Asp Glu
-40 -35 -30

Ala Arg Ser Leu Pro Pro Pro Lys Leu Thr Asp Pro Arg Leu Leu Tyr -25 -15 -10

Ile Gly Phe Leu Gly Tyr Cys Ser Gly Leu Ile Asp Asn Leu Ile Arg
-5 1 5

Arg Arg Pro Ile Ala Thr Ala Gly Leu His Arg Gln Leu Leu Tyr Ile 10 15 20

Xaa Ala Gly

25

233

- (2) INFORMATION FOR SEQ ID NO: 293:
 - (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 48 amino acids
 - (B) TYPE: AMINO ACID
 - (D) TOPOLOGY: LINEAR
 - (ii) MOLECULE TYPE: PROTEIN
 - (vi) ORIGINAL SOURCE:
 - (A) ORGANISM: Homo Sapiens
 - (F) TISSUE TYPE: Thyroid
 - (ix) FEATURE:
 - (A) NAME/KEY: sig_peptide
 - (B) LOCATION: -24..-1
 - (C) IDENTIFICATION METHOD: Von Heijne matrix
 - (D) OTHER INFORMATION: score 5.2

seg VGGLILWLSVGSS/GD

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 293:

Met Pro Pro Gly Pro Trp Glu Ser Cys Phe Trp Val Gly Gly Leu Ile

Leu Trp Leu Ser Val Gly Ser Ser Gly Asp Ala Pro Pro Thr Pro Gln

Pro Lys Cys Ala Asp Phe Gln Ser Ala Asn Leu Phe Glu Gly Thr Arg 10

- (2) INFORMATION FOR SEQ ID NO: 294:
 - (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 36 amino acids
 - (B) TYPE: AMINO ACID
 - (D) TOPOLOGY: LINEAR
 - (ii) MOLECULE TYPE: PROTEIN
 - (vi) ORIGINAL SOURCE:
 - (A) ORGANISM: Homo Sapiens
 - (F) TISSUE TYPE: Lung
 - (ix) FEATURE:
 - (A) NAME/KEY: sig_peptide
 - (B) LOCATION: -14..-1
 - (C) IDENTIFICATION METHOD: Von Heijne matrix
 - (D) OTHER INFORMATION: score 5

seq CARALLLACSSRG/RH

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 294:

Met Cys Ala Arg Ala Leu Leu Leu Ala Cys Ser Ser Arg Gly Arg His -10

Arg Leu Ala Cys Gln Cys Ser Thr Ser Ala Thr Pro Ser Trp Ala Ala

15

10

Ala Ser Trp Gly

(2) INFORMATION FOR SEQ ID NO: 295:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 108 amino acids
 - (B) TYPE: AMINO ACID
 - (D) TOPOLOGY: LINEAR
- (ii) MOLECULE TYPE: PROTEIN
- (vi) ORIGINAL SOURCE:
 - (A) ORGANISM: Homo Sapiens
 - (F) TISSUE TYPE: Lung (cells)
- (ix) FEATURE:
 - (A) NAME/KEY: sig_peptide
 - (B) LOCATION: -46..-1
 - (C) IDENTIFICATION METHOD: Von Heijne matrix
 - (D) OTHER INFORMATION: score 4.7

seq IICCVFLLLAIVG/YV

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 295:

Met Gly Asp Glu Arg Pro His Tyr Tyr Gly Lys His Gly Thr Pro Gln

Lys Tyr Asp Pro Thr Phe Lys Gly Pro Ile Tyr Asn Arg Gly Cys Thr
-30 -25 -20 -15

Asp Ile Ile Cys Cys Val Phe Leu Leu Ala Ile Val Gly Tyr Val
-10 -5 1

Ala Val Gly Ile Ile Ala Trp Thr His Gly Asp Pro Arg Lys Val Ile
5 10 15

Tyr Pro Thr Asp Ser Arg Gly Glu Phe Cys Gly Gln Lys Gly Thr Lys 20 25 30

Asn Glu Asn Lys Pro Tyr Leu Phe Tyr Phe Asn Ile Val Lys Cys Ala 35 40 45 50

Ser Pro Leu Val Leu Leu Glu Phe Gln Cys Pro Thr 55 60

(2) INFORMATION FOR SEQ ID NO: 296:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 81 amino acids
 - (B) TYPE: AMINO ACID
 - (D) TOPOLOGY: LINEAR
- (ii) MOLECULE TYPE: PROTEIN

(vi) ORIGINAL SOURCE:

- (A) ORGANISM: Homo Sapiens
- (F) TISSUE TYPE: Lung (cells)
- (ix) FEATURE:
 - (A) NAME/KEY: sig_peptide
 - (B) LOCATION: -16..-1
 - (C) IDENTIFICATION METHOD: Von Heijne matrix
 - (D) OTHER INFORMATION: score 4.5

seq RLLLRRFLASVIS/RK

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 296:

Met Ala Gln Arg Leu Leu Arg Arg Phe Leu Ala Ser Val Ile Ser -15 -5

Arg Lys Pro Ser Gln Gly Gln Trp Pro Pro Leu Thr Ser Arg Ala Leu
1 5 10 15

Gln Thr Pro Gln Cys Ser Pro Gly Gly Leu Thr Val Thr Pro Asn Pro 20 25 30

Ala Arg Thr Ile Tyr Thr Thr Arg Ile Ser Leu Thr Thr Phe Asn Ile 35 40 45

Gln Asp Gly Pro Asp Phe Gln Asp Arg Val Val Asn Ser Glu Thr Pro 50 55 60

Ala 65

- (2) INFORMATION FOR SEQ ID NO: 297:
 - (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 90 amino acids
 - (B) TYPE: AMINO ACID
 - (D) TOPOLOGY: LINEAR
 - (ii) MOLECULE TYPE: PROTEIN
 - (vi) ORIGINAL SOURCE:
 - (A) ORGANISM: Homo Sapiens
 - (F) TISSUE TYPE: Lung (cells)
 - (ix) FEATURE:
 - (A) NAME/KEY: sig_peptide
 - (B) LOCATION: -23..-1
 - (C) IDENTIFICATION METHOD: Von Heijne matrix
 - (D) OTHER INFORMATION: score 4.4

seq QFILLGTTSVVTA/AL

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 297:

Met Glu Ser Gly Gly Arg Pro Ser Leu Cys Gln Phe Ile Leu Leu Gly

236

Thr Thr Ser Val Val Thr Ala Ala Leu Tyr Ser Val Tyr Arg Gln Lys
-5 1 5

Ala Arg Val Ser Gln Glu Leu Lys Gly Ala Lys Lys Val His Leu Gly 10 25

Glu Asp Leu Lys Ser Ile Leu Ser Glu Ala Pro Gly Lys Cys Val Pro 30 35 40

Xaa Ala Val Ile Glu Gly Ala Val Arg Ser Val Lys Glu Thr Leu Asn 45 50 55

Ser Gln Phe Val Glu Asn Cys Lys Gly Xaa 60 65

- (2) INFORMATION FOR SEQ ID NO: 298:
 - (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 73 amino acids
 - (B) TYPE: AMINO ACID
 - (D) TOPOLOGY: LINEAR
 - (ii) MOLECULE TYPE: PROTEIN
 - (vi) ORIGINAL SOURCE:
 - (A) ORGANISM: Homo Sapiens
 - (F) TISSUE TYPE: Liver
 - (ix) FEATURE:
 - (A) NAME/KEY: sig_peptide
 - (B) LOCATION: -24..-1
 - (C) IDENTIFICATION METHOD: Von Heijne matrix
 - (D) OTHER INFORMATION: score 16.4

seq LLLLLLASLTSG/SV

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 298:

Met Ala Leu Ser Ser Gln Ile Tro Ala Ala Cys Leu Leu Leu Leu -20 -15 -10

Leu Leu Ala Ser Leu Thr Ser Gly Ser Val Phe Pro Gln Gln Thr Gly
-5 1 5

Gln Leu Ala Glu Leu Gln Pro Gln Asp Arg Ala Gly Ala Arg Ala Ser 10 15 20

Trp Met Pro Met Phe Gln Arg Arg Arg Arg Arg Asp Thr His Phe Pro 25 30 35 40

Ile Cys Ile Phe Cys Cys Gly Pro Gly

- (2) INFORMATION FOR SEQ ID NO: 299:
 - (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 79 amino acids

237

- (B) TYPE: AMINO ACID
 (D) TOPOLOGY: LINEAR
- (ii) MOLECULE TYPE: PROTEIN
- (vi) ORIGINAL SOURCE:
 - (A) ORGANISM: Homo Sapiens
 - (F) TISSUE TYPE: Large intestine
- (ix) FEATURE:
 - (A) NAME/KEY: sig_peptide
 - (B) LOCATION: -23..-1
 - (C) IDENTIFICATION METHOD: Von Heijne matrix
 - (D) OTHER INFORMATION: score 14.4

seq LGLLLFLLPGSLG/AE

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 299:

Met Gly Val Pro Arg Pro Gln Pro Trp Ala Leu Gly Leu Leu Phe
-20 -15 -10

Leu Leu Pro Gly Ser Leu Gly Ala Glu Ser His Leu Ser Leu Leu Tyr
-5 1 5

His Leu Thr Ala Val Ser Ser Pro Ala Pro Gly Thr Pro Ala Phe Trp 10 15 20 25

Val Ser Gly Trp Leu Gly Pro Gln Gln Tyr Leu Ser Tyr Asn Ser Leu
30 35 40

Arg Gly Glu Ala Xaa Xaa Val Glu Leu Gly Ser Gly Lys Thr Arg
45 50 55

- (2) INFORMATION FOR SEQ ID NO: 300:
 - (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 51 amino acids
 - (B) TYPE: AMINO ACID
 - (D) TOPOLOGY: LINEAR
 - (ii) MOLECULE TYPE: PROTEIN
 - (vi) ORIGINAL SOURCE:
 - (A) ORGANISM: Homo Sapiens
 - (F) TISSUE TYPE: Liver
 - (ix) FEATURE:
 - (A) NAME/KEY: sig_peptide
 - (B) LOCATION: -18..-1
 - (C) IDENTIFICATION METHOD: Von Heijne matrix
 - (D) OTHER INFORMATION: score 10.3

seq SLLLSVLLAQVWL/VP

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 300:

Met Lys Val Val Pro Ser Leu Leu Ser Val Leu Leu Ala Gln Val
-15 -10 -5

Trp Leu Val Pro Gly Leu Ala Pro Ser Pro Gln Ser Pro Glu Thr Pro $1 \hspace{1cm} 5 \hspace{1cm} 10$

Ala Pro Gln Asn Gln Thr Ser Arg Val Val Gln Ala Pro Arg Glu Glu
15 20 25 30

Glu Glu Trp

(2) INFORMATION FOR SEQ ID NO: 301:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 195 amino acids
 - (B) TYPE: AMINO ACID
 - (D) TOPOLOGY: LINEAR
- (ii) MOLECULE TYPE: PROTEIN
- (vi) ORIGINAL SOURCE:
 - (A) ORGANISM: Homo Sapiens
 - (F) TISSUE TYPE: Pancreas
- (ix) FEATURE:
 - (A) NAME/KEY: sig_peptide
 - (B) LOCATION: -16..-1
 - (C) IDENTIFICATION METHOD: Von Heijne matrix
 - (D) OTHER INFORMATION: score 9.5

seq ITVLAALLACASS/CG

- (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 301:
- Met Leu Ser Ile Thr Val Leu Ala Ala Leu Leu Ala Cys Ala Ser Ser -15 -5
- Cys Gly Val Pro Ser Phe Pro Pro Asn Leu Ser Ala Arg Xaa Val Gly
 1 5 10 15
- Gly Glu Asp Ala Arg Pro His Ser Trp Pro Trp Gln Ile Ser Leu Gln
 20 25 30
- Tyr Leu Lys Asn Asp Thr Trp Arg His Thr Cys Gly Gly Thr Leu Ile 35 40 45
- Ala Ser Asn Phe Xaa Leu Thr Ala Ala His Cys Ile Ser Asn Thr Arg 50 55 60
- Thr Tyr Arg Val Ala Val Gly Lys Asn Asn Leu Glu Val Glu Asp Glu 65 70 75 80
- Glu Gly Ser Leu Phe Val Gly Val Asp Thr Ile His Val His Lys Xaa 85 90 95
- Xaa Asn Ala Xaa Leu Leu Arg Asn Asp Ile Ala Leu Ile Lys Leu Ala 100 105 110
- Glu His Val Glu Leu Ser Asp Thr Ile Gln Val Ala Cys Xaa Pro Glu 115 120 125

239

Lys Asp Ser Leu Leu Pro Lys Asp Tyr Pro Cys Tyr Val Ser Arg Leu 130 135 140

Xaa Pro Pro Xaa Gly Gly Xaa Ser Gly Gly Xaa Leu Asn Cys Gln Leu 145 150 155 160

Glu Asn Gly Ser Trp Glu Val Phe Gly Xaa Val Ser Phe Gly Ser Arg 165 170 175

Arg Gly Cys

(2) INFORMATION FOR SEQ ID NO: 302:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 123 amino acids
 - (B) TYPE: AMINO ACID
 - (D) TOPOLOGY: LINEAR
- (ii) MOLECULE TYPE: PROTEIN
- (vi) ORIGINAL SOURCE:
 - (A) ORGANISM: Homo Sapiens
 - (F) TISSUE TYPE: Pancreas
- (ix) FEATURE:
 - (A) NAME/KEY: sig_peptide
 - (B) LOCATION: -16..-1
 - (C) IDENTIFICATION METHOD: Von Heijne matrix
 - (D) OTHER INFORMATION: score 9.5 seq ITVLAALLACASS/CG
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 302:

Met Leu Gly Ile Thr Val Leu Ala Ala Leu Leu Ala Cys Ala Ser Ser -15 -5

Cys Gly Val Pro Ser Phe Pro Pro Asn Leu Ser Ala Arg Val Val Gly
1 5 10 15

Gly Glu Asp Ala Arg Pro His Xaa Trp Pro Trp Gln Ile Ser Leu Gln 20 25 30

Tyr Leu Lys Asn Asp Thr Trp Arg His Thr Cys Gly Gly Thr Leu Ile 35 40 45

Ala Ser Asn Phe Val Leu Thr Ala Ala His Cys Ile Ser Xaa Thr Arg 50 60

Thr Tyr Arg Val Ala Val Gly Lys Asn Asn Leu Glu Val Glu Asp Glu 65 70 75 80

Glu Gly Ser Leu Xaa Val Gly Val Asp Thr Ile His Val His Arg Arg 85 90 95

Trp Asn Ala Leu Leu Leu Arg Asn Asp Ile Ala

240

(2) INFORMATION FOR SEQ ID NO: 303:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 52 amino acids
- (B) TYPE: AMINO ACID
- (D) TOPOLOGY: LINEAR
- (ii) MOLECULE TYPE: PROTEIN
- (vi) ORIGINAL SOURCE:
 - (A) ORGANISM: Homo Sapiens
 - (F) TISSUE TYPE: Colon
- (ix) FEATURE:
 - (A) NAME/KEY: sig_peptide
 - (B) LOCATION: -38..-1
 - (C) IDENTIFICATION METHOD: Von Heijne matrix
 - (D) OTHER INFORMATION: score 8.8

seq LLLPLLSLPVTTP/WT

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 303:

Met Ala Gly Asn Gly Glu Ser Glu Pro Asp Arg Leu His Leu Leu Thr

Gly His Arg Val Lys Gly Glu Phe Gln Leu Leu Leu Pro Leu Leu Ser

Leu Pro Val Thr Thr Pro Trp Thr Asn Pro Glu Glu Gly Thr Phe Ser

Arg Ser His Gly

- (2) INFORMATION FOR SEQ ID NO: 304:
 - (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 90 amino acids
 - (B) TYPE: AMINO ACID
 - (D) TOPOLOGY: LINEAR
 - (ii) MOLECULE TYPE: PROTEIN
 - (vi) ORIGINAL SOURCE:
 - (A) ORGANISM: Homo Sapiens
 - (F) TISSUE TYPE: Large intestine
 - (ix) FEATURE:
 - (A) NAME/KEY: sig peptide
 - (B) LOCATION: -14..-1
 - (C) IDENTIFICATION METHOD: Von Heijne matrix
 - (D) OTHER INFORMATION: score 8.7

seq LWWLVLLLLPTLK/SV

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 304:

Met Leu Trp Trp Leu Val Leu Leu Leu Leu Pro Thr Leu Lys Ser Val

241

Phe Cys Ser Leu Val Thr Ser Leu Tyr Leu Pro Asn Thr Glu Asp Leu 5 10 15

Ser Leu Trp Leu Trp Pro Lys Pro Asp Leu His Ser Gly Thr Arg Thr 20 25 30

Glu Val Ser Thr His Thr Val Pro Ser Lys Pro Gly Thr Ala Ser Pro 35 40 45 50

Cys Trp Pro Leu Ala Gly Ala Val Pro Ser Pro Thr Val Ser Arg Leu
55 60 65

Glu Ala Leu Thr Arg Ala Val Gln Val Ala
70
75

- (2) INFORMATION FOR SEQ ID NO: 305:
 - (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 53 amino acids
 - (B) TYPE: AMINO ACID
 - (D) TOPOLOGY: LINEAR
 - (ii) MOLECULE TYPE: PROTEIN
 - (vi) ORIGINAL SOURCE:
 - (A) ORGANISM: Homo Sapiens
 - (F) TISSUE TYPE: Liver
 - (ix) FEATURE:
 - (A) NAME/KEY: sig_peptide
 - (B) LOCATION: -29..-1
 - (C) IDENTIFICATION METHOD: Von Heijne matrix
 - (D) OTHER INFORMATION: score 8.5

seq LLGLLMAACFTFC/LS

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 305:

Met Ala Pro Gln Ser Leu Pro Ser Ser Arg Met Ala Pro Leu Gly Met
-25 -20 -15

Leu Leu Gly Leu Leu Met Ala Ala Cys Phe Thr Phe Cys Leu Ser His -10 -5 1

Gln Asn Leu Lys Glu Tyr Ala Leu Thr Asn Pro Xaa Lys Xaa Ser Thr 5 10

Lys Glu Thr Glu Gly 20

- (2) INFORMATION FOR SEQ ID NO: 306:
 - (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 70 amino acids
 - (B) TYPE: AMINO ACID

PCT/IB98/01233 242

- (D) TOPOLOGY: LINEAR
- (ii) MOLECULE TYPE: PROTEIN
- (vi) ORIGINAL SOURCE:
 - (A) ORGANISM: Homo Sapiens
 - (F) TISSUE TYPE: Lung
- (ix) FEATURE:
 - (A) NAME/KEY: sig_peptide
 - (B) LOCATION: -23..-1
 - (C) IDENTIFICATION METHOD: Von Heijne matrix
 - (D) OTHER INFORMATION: score 8.1

seq LLLVTVSSNLAIA/IK

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 306:

Met Met Leu His Ser Ala Leu Gly Leu Cys Leu Leu Leu Val Thr Val -15

Ser Ser Asn Leu Ala Ile Ala Ile Lys Lys Glu Lys Arg Pro Pro Gln

Thr Leu Ser Arg Gly Trp Gly Asp Asp Ile Thr Trp Val Gln Thr Tyr

Glu Glu Gly Leu Phe Tyr Ala Gln Lys Ser Lys Lys Pro Leu Met Val

Ile His His Leu Asp Gly

- (2) INFORMATION FOR SEQ ID NO: 307:
 - (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 86 amino acids
 - (B) TYPE: AMINO ACID
 - (D) TOPOLOGY: LINEAR
 - (ii) MOLECULE TYPE: PROTEIN
 - (vi) ORIGINAL SOURCE:
 - (A) ORGANISM: Homo Sapiens
 - (F) TISSUE TYPE: Liver
 - (ix) FEATURE:
 - (A) NAME/KEY: sig peptide
 - (B) LOCATION: -26..-1
 - (C) IDENTIFICATION METHOD: Von Heijne matrix
 - (D) OTHER INFORMATION: score 7.6

seq LITLCLVCIVANA/LL

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 307:

Met Cys Thr Gly Lys Cys Ala Arg Cys Val Gly Leu Ser Leu Ile Thr -20

243

Leu Cys Leu Val Cys Ile Val Ala Asn Ala Leu Leu Val Pro Asn -10 -5 1 5

Gly Glu Thr Ser Trp Thr Asn Thr Asn His Leu Ser Leu Gln Val Trp 10 15 20

Leu Met Gly Gly Phe Ile Gly Gly Gly Leu Met Val Leu Cys Pro Gly
25 30 35

Ile Ala Ala Val Arg Ala Gly Gly Lys Xaa Cys Cys Gly Ala Gly Cys 40 50

Cys Gly Asn Arg Leu Arg
55 60

- (2) INFORMATION FOR SEQ ID NO: 308:
 - (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 36 amino acids
 - (B) TYPE: AMINO ACID
 - (D) TOPOLOGY: LINEAR
 - (ii) MOLECULE TYPE: PROTEIN
 - (vi) ORIGINAL SOURCE:
 - (A) ORGANISM: Homo Sapiens
 - (F) TISSUE TYPE: Lung (cells)
 - (ix) FEATURE:
 - (A) NAME/KEY: sig_peptide
 - (B) LOCATION: -23..-1
 - (C) IDENTIFICATION METHOD: Von Heijne matrix
 - (D) OTHER INFORMATION: score 7.5

seq LVLLLTLPLHLMA/LL

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 308:

Met Asp Ile Leu Val Pro Leu Leu Gln Leu Leu Val Leu Leu Thr -20 -15 -10

Leu Pro Leu His Leu Met Ala Leu Leu Gly Cys Trp Gln Pro Leu Cys
-5 5

Lys Ser Phe Gly

- (2) INFORMATION FOR SEQ ID NO: 309:
 - (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 23 amino acids
 - (B) TYPE: AMINO ACID
 - (D) TOPOLOGY: LINEAR
 - (ii) MOLECULE TYPE: PROTEIN
 - (vi) ORIGINAL SOURCE:

244

(A) ORGANISM: Homo Sapiens
(F) TISSUE TYPE: Liver

(ix) FEATURE:

(A) NAME/KEY: sig_peptide

(B) LOCATION: -15..-1

- (C) IDENTIFICATION METHOD: Von Heijne matrix
- (D) OTHER INFORMATION: score 7.5

seq FLVLFSFFNIALC/AP

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 309:

Met Pro Phe Leu Val Leu Phe Ser Phe Phe Asn Ile Ala Leu Cys Ala
-15 -5 1

Pro Arg Lys Phe Ala Arg Lys

- (2) INFORMATION FOR SEQ ID NO: 310:
 - (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 108 amino acids
 - (B) TYPE: AMINO ACID
 - (D) TOPOLOGY: LINEAR
 - (ii) MOLECULE TYPE: PROTEIN
 - (vi) ORIGINAL SOURCE:
 - (A) ORGANISM: Homo Sapiens
 - (F) TISSUE TYPE: Pancreas
 - (ix) FEATURE:
 - (A) NAME/KEY: sig_peptide
 - (B) LOCATION: -19..-1
 - (C) IDENTIFICATION METHOD: Von Heijne matrix
 - (D) OTHER INFORMATION: score 6.8

seq AIVALAVCAALHA/SE

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 310:

Met Gln Gln Arg Gly Leu Ala Ile Val Ala Leu Ala Val Cys Ala Ala
-15
-10
-5

Leu His Ala Ser Glu Ala Ile Leu Pro Ile Ala Ser Ser Cys Cys Thr $1 \hspace{1cm} 5 \hspace{1cm} 10$

Glu Val Ser His His Ile Ser Arg Arg Leu Leu Glu Arg Val Asn Met 15 20 25

Cys Arg Ile Gln Arg Ala Asp Gly Asp Cys Asp Leu Ala Ala Val Ile 30 40 45

Leu His Val Lys Arg Arg Ile Cys Val Ser Pro His Asn His Thr
50 55 60

Vai Lys Gln Trp Met Lys Val Gln Ala Ala Lys Lys Asn Gly Lys Gly
65 70 75

Asn Val Cys His Arg Lys Lys His His Gly Lys Arg 80 85

(2) INFORMATION FOR SEQ ID NO: 311:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 108 amino acids
- (B) TYPE: AMINO ACID
- (D) TOPOLOGY: LINEAR
- (ii) MOLECULE TYPE: PROTEIN
- (vi) ORIGINAL SOURCE:
 - (A) ORGANISM: Homo Sapiens
 - (F) TISSUE TYPE: Lung
- (ix) FEATURE:
 - (A) NAME/KEY: sig_peptide
 - (B) LOCATION: -20..-1
 - (C) IDENTIFICATION METHOD: Von Heijne matrix
 - (D) OTHER INFORMATION: score 6.8

seq GLLWMLFVSELRA/AT

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 311:

Met Arg Lys Thr Arg Leu Trp Gly Leu Leu Trp Met Leu Phe Val Ser
-20 -15 -10 -5

Glu Leu Arg Ala Ala Thr Lys Leu Thr Glu Glu Lys Tyr Glu Leu Lys

1 5 10

Glu Gly Gln Thr Leu Asp Val Lys Cys Asp Tyr Thr Leu Glu Lys Phe 15 20 25

Ala Ser Ser Gln Lys Ala Trp Gln Ile Ile Arg Asp Gly Glu Met Pro $30 \hspace{1cm} 35 \hspace{1cm} 40$

Lys Thr Leu Ala Cys Thr Glu Arg Pro Ser Lys Asn Ser His Pro Val 45 50 55 60

Gln Val Gly Arg Ile Ile Leu Glu Asp Tyr His Asp His Gly Leu Leu 65 70 75

Arg Val Arg Met Val Asn Leu Gln Val Xaa Asp Ser 80 85

- (2) INFORMATION FOR SEQ ID NO: 312:
 - (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 64 amino acids
 - (B) TYPE: AMINO ACID
 - (D) TOPOLOGY: LINEAR
 - (ii) MOLECULE TYPE: PROTEIN

- (vi) ORIGINAL SOURCE:
 - (A) ORGANISM: Homo Sapiens
 - (F) TISSUE TYPE: Thyroid
- (ix) FEATURE:
 - (A) NAME/KEY: sig_peptide
 - (B) LOCATION: -22..-1
 - (C) IDENTIFICATION METHOD: Von Heijne matrix
 - (D) OTHER INFORMATION: score 6.8

seq VSLVLLMPGPCDG/LF

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 312:

Met Val Gly Ala Met Trp Lys Val Ile Val Ser Leu Val Leu Met
-20 -15 -10

Pro Gly Pro Cys Asp Gly Leu Phe Arg Ser Leu Tyr Arg Ser Val Xaa -5 5 10

Met Pro Pro Lys Gly Asp Ser Gly Gln Pro Leu Phe Leu Thr Pro Tyr
15 20 25

Ile Glu Ala Gly Lys Ile Gln Lys Gly Arg Glu Xaa Xaa Leu Val Gly
30 35 40

- (2) INFORMATION FOR SEQ ID NO: 313:
 - (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 61 amino acids
 - (B) TYPE: AMINO ACID
 - (D) TOPOLOGY: LINEAR
 - (ii) MOLECULE TYPE: PROTEIN
 - (vi) ORIGINAL SOURCE:
 - (A) ORGANISM: Homo Sapiens
 - (F) TISSUE TYPE: Lung (cells)
 - (ix) FEATURE:
 - (A) NAME/KEY: sig_peptide
 - (B) LOCATION: -40..-1
 - (C) IDENTIFICATION METHOD: Von Heijne matrix
 - (D) OTHER INFORMATION: score 6.1

seq ICIGILVLPFIRC/CC

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 313:

Met Ile His Leu Arg Ile Ile Gln Arg Cys Tyr Met Ala Gly Leu Glu
-40 -35 -30 -25

Asr Lys Lys Asr Val Val Phe Glu Ala Lys Gln Ile Cys Ile Gly Ile
-20 -15 -10

Leu Val Leu Pro Phe Ile Arg Cys Cys Leu Val Gln Ile Thr Phe
-5 1 5

Ser Leu Ser Leu His Phe Leu Ile Tyr Asn Met Arg Arg

10 15 20

- (2) INFORMATION FOR SEQ ID NO: 314:
 - (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 98 amino acids
 - (B) TYPE: AMINO ACID
 - (D) TOPOLOGY: LINEAR
 - (ii) MOLECULE TYPE: PROTEIN
 - (vi) ORIGINAL SOURCE:
 - (A) ORGANISM: Homo Sapiens
 - (F) TISSUE TYPE: Lung (cells)
 - (ix) FEATURE:
 - (A) NAME/KEY: sig_peptide
 - (B) LOCATION: -22..-1
 - (C) IDENTIFICATION METHOD: Von Heijne matrix
 - (D) OTHER INFORMATION: score 5.9

seq LIYILWQLTGSAA/SG

- (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 314:
- Met Ala Gly Ser Pro Thr Cys Leu Thr Leu Ile Tyr Ile Leu Trp Gln
 -20 -15 -10
- Leu Thr Gly Ser Ala Ala Ser Gly Pro Val Lys Glu Leu Val Gly Ser
 -5 5 10
- Val Gly Gly Ala Val Thr Phe Pro Leu Lys Ser Lys Val Lys Gln Val
 15 20 25
- Asp Ser Ile Val Trp Thr Phe Asn Thr Thr Pro Leu Val Thr Ile Gln 30 35 40
- Pro Glu Gly Gly Thr Ile Ile Val Thr Gln Asn Arg Asn Arg Glu Arg
 45 50 55
- Val Asp Phe Pro Asp Gly Gly Tyr Ser Leu Lys Leu Ser Lys Leu Lys
 60 70

Lys Gly 75

- (2) INFORMATION FOR SEQ ID NO: 315:
 - (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 113 amino acids
 - (B) TYPE: AMINO ACID
 - (D) TOPOLOGY: LINEAR
 - (ii) MOLECULE TYPE: PROTEIN
 - (vi) ORIGINAL SOURCE:
 - (A) ORGANISM: Homo Sapiens

- (F) TISSUE TYPE: Thyroid
- (ix) FEATURE:
 - (A) NAME/KEY: sig peptide
 - (B) LOCATION: -59..-1
 - (C) IDENTIFICATION METHOD: Von Heijne matrix
 - (D) OTHER INFORMATION: score 5.7

seq ALLDLCAAPXGWL/QV

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 315:

Met Gly Lys Lys Gly Lys Val Gly Lys Ser Arg Arg Asp Lys Phe Tyr
-55 -50 -45

His Leu Ala Lys Glu Thr Gly Tyr Arg Ser Arg Ser Ala Phe Lys Leu -40 -35 -30

Ile Gln Leu Asn Arg Arg Phe Gln Phe Leu Gln Lys Ala Arg Ala Leu
-25 -20 -15

Leu Asp Leu Cys Ala Ala Pro Xaa Gly Trp Leu Gln Val Ala Ala Lys
-10 -5 1 5

Phe Met Pro Val Ser Ser Leu Ile Val Gly Val Asp Leu Val Pro Ile 10 15 20

Lys Pro Leu Pro Asn Val Val Thr Leu Gln Glu Asp Ile Thr Thr Glu 25 30 35

Arg Cys Xaa Gln Arg His Trp Thr Ser Ala Ser Ala Leu Glu Arg Lys 40 45 50

Met

- (2) INFORMATION FOR SEQ ID NO: 316:
 - (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 75 amino acids
 - (B) TYPE: AMINO ACID
 - (D) TOPOLOGY: LINEAR
 - (ii) MOLECULE TYPE: PROTEIN
 - (vi) ORIGINAL SOURCE:
 - (A) ORGANISM: Homo Sapiens
 - (D) DEVELOPMENTAL STAGE: Fetal
 - (F) TISSUE TYPE: liver
 - (ix) FEATURE:
 - (A) NAME/KEY: sig_peptide
 - (B) LOCATION: -32..-1
 - (C) IDENTIFICATION METHOD: Von Heijne matrix
 - (D) OTHER INFORMATION: score 5.7

seg FGLVXVGTALALA/RK

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 316:

249

Met Pro Leu Ser Asp Phe Ile Leu Ala Leu Lys Asp Asn Pro Tyr Phe -30 -25 -20

Gly Ala Gly Phe Gly Leu Val Xaa Val Gly Thr Ala Leu Ala Leu Ala -15 -5

Arg Lys Gly Val Gln Leu Gly Leu Val Ala Phe Arg Arg His Tyr Met

1 5 10 15

Ile Thr Leu Glu Val Pro Ala Arg Asp Arg Ser Tyr Xaa Trp Leu Leu 20 25 30

Ser Trp Leu Thr Arg His Ser Thr Arg Thr Gly
35 40

(2) INFORMATION FOR SEQ ID NO: 317:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 93 amino acids
 - (B) TYPE: AMINO ACID
 - (D) TOPOLOGY: LINEAR
- (ii) MOLECULE TYPE: PROTEIN
- (vi) ORIGINAL SOURCE:
 - (A) ORGANISM: Homo Sapiens
 - (F) TISSUE TYPE: Large intestine
- (ix) FEATURE:
 - (A) NAME/KEY: sig_peptide
 - (B) LOCATION: -19..-1
 - (C) IDENTIFICATION METHOD: Von Heijne matrix
 - (D) OTHER INFORMATION: score 5.6

seq WVFLVAIIKGVQC/QA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 317:

Met Glu Phe Gly Leu Ser Trp Val Phe Leu Val Ala Ile Ile Lys Gly
-15 -10 -5

Val Gln Cys Gln Ala Gln Leu Glu Glu Ser Gly Gly Gly Leu Val Gln
1 5 10

Pro Gly Gly Ser Leu Arg Leu Ser Cys Arg Gly Ser Gly Phe Thr Leu 15 20 25

Ser Asp His Tyr Met Ser Trp Ile Arg Gln Ser Pro Gly Lys Gly Xaa 30 40 45

Xaa Trp Val Ala Tyr Ile Ser Tyr Ser Gly Ser Thr Ile Tyr Tyr Gly
50 55 60

Asp Ser Val Asp Gly Arg Phe Thr Ile Ser Arg Asp Asn 65 . 70

(2) INFORMATION FOR SEQ ID NO: 318:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 44 amino acids

(B) TYPE: AMINO ACID (D) TOPOLOGY: LINEAR

- (ii) MOLECULE TYPE: PROTEIN
- (vi) ORIGINAL SOURCE:
 - (A) ORGANISM: Homo Sapiens
 - (F) TISSUE TYPE: Lung (cells)
- (ix) FEATURE:
 - (A) NAME/KEY: sig_peptide
 - (B) LOCATION: -21..-1
 - (C) IDENTIFICATION METHOD: Von Heijne matrix
 - (D) OTHER INFORMATION: score 5.3

seq SLFSSLPIFLTWA/HI

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 318:

Met Ile Leu Arg Lys Arg Ser Cys Ser Leu Phe Ser Ser Leu Pro Ile
-20 -15 -10

Phe Leu Thr Trp Ala His Ile Lys Arg Val Pro Leu Leu Xaa Thr Ser -5 1 5 10

Leu His Thr Ala His Asn Gly His Pro His Tyr Gly
15 20

- (2) INFORMATION FOR SEQ ID NO: 319:
 - (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 30 amino acids
 - (B) TYPE: AMINO ACID
 - (D) TOPOLOGY: LINEAR
 - (ii) MOLECULE TYPE: PROTEIN
 - (vi) ORIGINAL SOURCE:
 - (A) ORGANISM: Homo Sapiens
 - (F) TISSUE TYPE: Colon
 - (ix) FEATURE:
 - (A) NAME/KEY: sig_peptide
 - (B) LOCATION: -16..-1
 - (C) IDENTIFICATION METHOD: Von Heijne matrix
 - (D) OTHER INFORMATION: score 5

seq GLMFVKLVNPCSG/EG

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 319:

Met Lys Asn Gly Leu Met Phe Val Lys Leu Val Asn Pro Cys Ser Gly
-15 -5

Glu Gly Ala Ile Tyr Leu Phe Asn Met Cys Leu Gln Gln Arg
1 5 10

251

- (2) INFORMATION FOR SEQ ID NO: 320:
 - (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 39 amino acids
 - (B) TYPE: AMINO ACID
 - (D) TOPOLOGY: LINEAR
 - (ii) MOLECULE TYPE: PROTEIN
 - (vi) ORIGINAL SOURCE:
 - (A) ORGANISM: Homo Sapiens
 - (F) TISSUE TYPE: Lung (cells)
 - (ix) FEATURE:
 - (A) NAME/KEY: sig_peptide
 - (B) LOCATION: -15..-1
 - (C) IDENTIFICATION METHOD: Von Heijne matrix
 - (D) OTHER INFORMATION: score 4.9

seq AVVFVFSLLDCCA/LI

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 320:

Met Glu Ala Val Val Phe Val Phe Ser Leu Leu Asp Cys Cys Ala Leu
-15 -5 1

Ile Phe Leu Ser Val Tyr Phe Ile Ile Thr Leu Ser Asp Leu Glu Cys
5 10 15

Asp Tyr Ile Asn Ala Arg Ser 20

- (2) INFORMATION FOR SEQ ID NO: 321:
 - (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 62 amino acids
 - (B) TYPE: AMINO ACID
 - (D) TOPOLOGY: LINEAR
 - (ii) MOLECULE TYPE: PROTEIN
 - (vi) ORIGINAL SOURCE:
 - (A) ORGANISM: Homo Sapiens
 - (F) TISSUE TYPE: Pancreas
 - (ix) FEATURE:
 - (A) NAME/KEY: sig_peptide
 - (B) LOCATION: -56..-1
 - (C) IDENTIFICATION METHOD: Von Heijne matrix
 - (D) OTHER INFORMATION: score 4.7

seq FACVPGASXTTLA/FP

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 321:

Met Thr Gly Phe Leu Leu Pro Pro Ala Ser Arg Gly Thr Arg Arg Ser

252

-55 **-**50 **-**45

Cys Ser Arg Ser Arg Lys Arg Gln Thr Arg Arg Arg Arg Asn Pro Ser -40 -35 -30 -25

Ser Phe Val Ala Ser Cys Pro Thr Leu Leu Pro Phe Ala Cys Val Pro -20 -15 -10

Gly Ala Ser Xaa Thr Thr Leu Ala Phe Pro Pro Val Val Leu
-5 1 5

- (2) INFORMATION FOR SEQ ID NO: 322:
 - (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 60 amino acids
 - (B) TYPE: AMINO ACID(D) TOPOLOGY: LINEAR
 - (ii) MOLECULE TYPE: PROTEIN
 - (vi) ORIGINAL SOURCE:
 - (A) ORGANISM: Homo Sapiens
 - (F) TISSUE TYPE: Colon
 - (ix) FEATURE:
 - (A) NAME/KEY: sig_peptide
 - (B) LOCATION: -19..-1
 - (C) IDENTIFICATION METHOD: Von Heijne matrix
 - (D) OTHER INFORMATION: score 4.5 seq SVPLLTDAATVSG/AE
 - (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 322:

Met Cys Gly Asn Thr Met Ser Val Pro Leu Leu Thr Asp Ala Ala Thr -15 -10 -5

Val Ser Gly Ala Glu Arg Glu Thr Ala Ala Val Ile Phe Leu His Gly
1 5 10

Leu Gly Asp Thr Gly His Ser Trp Ala Asp Ala Leu Ser Thr Ile Arg
15 20 25

Leu Pro His Val Lys Tyr Ile Cys Pro His Ala Arg
30 40

- (2) INFORMATION FOR SEQ ID NO: 323:
 - (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 61 amino acids
 - (B) TYPE: AMINO ACID
 - (D) TOPOLOGY: LINEAR
 - (ii) MOLECULE TYPE: PROTEIN
 - (vi) ORIGINAL SOURCE:
 - (A) ORGANISM: Homo Sapiens

(F) TISSUE TYPE: Large intestine

- (ix) FEATURE:
 - (A) NAME/KEY: sig_peptide
 - (B) LOCATION: -43..-1
 - (C) IDENTIFICATION METHOD: Von Heijne matrix
 - (D) OTHER INFORMATION: score 4.3

seq SLWRLQWLKDASC/DP

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 323:

Met Xaa Xaa Xaa Glu Arg Arg Thr Ser Pro His Val Met Ala Asp
-40 -35 -30

Gln Ser Ser Thr Arg Asn Glu Asp Phe Leu Lys Lys Thr Trp Ser Leu
-25 -20 -15

Trp Arg Leu Gln Trp Leu Lys Asp Ala Ser Cys Asp Pro Tyr Pro Ala
-10 -5 1 5

Leu Pro Xaa Phe Trp Xaa Thr Glu Ala Lys Cys Glu Ser 10

- (2) INFORMATION FOR SEQ ID NO: 324:
 - (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 27 amino acids
 - (B) TYPE: AMINO ACID
 - (D) TOPOLOGY: LINEAR
 - (ii) MOLECULE TYPE: PROTEIN
 - (vi) ORIGINAL SOURCE:
 - (A) ORGANISM: Homo Sapiens
 - (F) TISSUE TYPE: Pancreas
 - (ix) FEATURE:
 - (A) NAME/KEY: sig_peptide
 - (B) LOCATION: -14..-1
 - (C) IDENTIFICATION METHOD: Von Heijne matrix
 - (D) OTHER INFORMATION: score 3.8 seq FLLLNCIVAVSQN/MG
 - (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 324:

Met Phe Leu Leu Asn Cys Ile Val Ala Val Ser Gln Asn Met Gly
-10 -5 1

Ile Gly Lys Asn Gly Asp Leu Pro Xaa Pro Gln
5 10

- (2) INFORMATION FOR SEQ ID NO: 325:
 - (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 60 amino acids

- (B) TYPE: AMINO ACID
- (D) TOPOLOGY: LINEAR
- (ii) MOLECULE TYPE: PROTEIN
- (vi) ORIGINAL SOURCE:
 - (A) ORGANISM: Homo Sapiens
 - (F) TISSUE TYPE: Lung
- (ix) FEATURE:
 - (A) NAME/KEY: sig_peptide
 - (B) LOCATION: -14..-1
 - (C) IDENTIFICATION METHOD: Von Heijne matrix
 - (D) OTHER INFORMATION: score 3.8

seq LLVSAAPLGFGQG/VW

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 325:

Met Leu Leu Val Ser Ala Ala Pro Leu Gly Phe Gly Gln Gly Val Trp
-10 -5

Asn Arg Ala Ser Gln Leu Gln Gln Gly Xaa Asp Pro Leu Gly Ala Gly
5 10 15

Arg Ser Trp Arg Gly Leu Cys Lys Leu Ser Gln Ala Leu Gly Ala Gly 20 25 30

Thr Gly Ser Gly Phe His Thr His Thr Arg Ala Pro 35 40 45

- (2) INFORMATION FOR SEQ ID NO: 326:
 - (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 36 amino acids
 - (B) TYPE: AMINO ACID
 - (D) TOPOLOGY: LINEAR
 - (ii) MOLECULE TYPE: PROTEIN
 - (vi) ORIGINAL SOURCE:
 - (A) ORGANISM: Homo Sapiens
 - (D) DEVELOPMENTAL STAGE: Fetal
 - (F) TISSUE TYPE: liver
 - (ix) FEATURE:
 - (A) NAME/KEY: sig_peptide
 - (B) LOCATION: -16..-1
 - (C) IDENTIFICATION METHOD: Von Heijne matrix
 - (D) OTHER INFORMATION: score 3.7

seq IALTLIPSMLSRA/AG

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 326:

Met Leu Arg Ile Ala Leu Thr Leu Ile Pro Ser Met Leu Ser Arg Ala
-15 -5

Ala Gly Trp Cys Trp Tyr Lys Glu Pro Thr Gln Gln Phe Ser Tyr Leu

255

1 5 10 15

Cys Leu Pro Gly

- (2) INFORMATION FOR SEQ ID NO: 327:
 - (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 41 amino acids
 - (B) TYPE: AMINO ACID
 - (D) TOPOLOGY: LINEAR
 - (ii) MOLECULE TYPE: PROTEIN
 - (vi) ORIGINAL SOURCE:
 - (A) ORGANISM: Homo Sapiens
 - (F) TISSUE TYPE: Lung
 - (ix) FEATURE:
 - (A) NAME/KEY: sig_peptide
 - (B) LOCATION: -19..-1
 - (C) IDENTIFICATION METHOD: Von Heijne matrix
 - (D) OTHER INFORMATION: score 3.7

seq LPGLRCSVPGVAA/RL

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 327:

Met Thr Leu Gly Gly Arg Leu Pro Gly Leu Arg Cys Ser Val Pro Gly
-15 -10 -5

Val Ala Ala Arg Leu Ser Thr Pro Pro Gln Val Arg Gln His Val Phe
1 5 10

Trp Ala Ala Ser Val Cys Xaa Xaa Thr 15 20

- (2) INFORMATION FOR SEQ ID NO: 328:
 - (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 67 amino acids
 - (B) TYPE: AMINO ACID
 - (D) TOPOLOGY: LINEAR
 - (ii) MOLECULE TYPE: PROTEIN
 - (vi) ORIGINAL SOURCE:
 - (A) ORGANISM: Homo Sapiens
 - (F) TISSUE TYPE: Liver
 - (ix) FEATURE:
 - (A) NAME/KEY: sig_peptide
 - (B) LOCATION: -14..-1
 - (C) IDENTIFICATION METHOD: Von Heijne matrix
 - (D) OTHER INFORMATION: score 3.6

seq AFTLXSLLQAALL/CV

256

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 328:

Met Ala Phe Thr Leu Xaa Ser Leu Leu Gln Ala Ala Leu Leu Cys Val

Asn Ala Ile Ala Val Leu His Glu Glu Arg Phe Leu Lys Asn Ile Gly
5 10 15

Trp Gly Thr Asp Gln Gly Ile Gly Gly Phe Gly Glu Glu Pro Gly Ile 20 25 30

Lys Ser Xaa Xaa Met Xaa Leu Ile Arg Ser Val Arg Thr Val Met Arg 35 40 45

Val Pro Leu

(2) INFORMATION FOR SEQ ID NO: 329:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 92 amino acids
 - (B) TYPE: AMINO ACID
 - (D) TOPOLOGY: LINEAR
- (ii) MOLECULE TYPE: PROTEIN
- (vi) ORIGINAL SOURCE:
 - (A) ORGANISM: Homo Sapiens
 - (F) TISSUE TYPE: Pancreas
- (ix) FEATURE:
 - (A) NAME/KEY: sig_peptide
 - (B) LOCATION: -21..-1
 - (C) IDENTIFICATION METHOD: Von Heijne matrix
 - (D) OTHER INFORMATION: score 3.5 seq LKVVFMVFASLXA/WY
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 329:

Met Arg Pro Leu Ala Gly Gly Leu Leu Lys Val Val Phe Met Val Phe -20 -15 -10

Ala Ser Leu Xaa Ala Trp Tyr Ser Gly Tyr Leu Leu Ala Xaa Xaa Ile
-5 10

Pro Asp Ala Pro Leu Ser Ser Ala Ala Tyr Ser Ile Arg Ser Ile Gly
15 20 25

Glu Arg Pro Val Leu Lys Ala Pro Val Pro Lys Arg Gln Lys Cys Asp 30 35 40

His Trp Thr Pro Cys Pro Ser Xaa Xaa Tyr Ala Tyr Arg Leu Leu Ser
45 50 55

Gly Gly Gly Arg Ser Lys Tyr Ala Lys Ile Cys Phe 60 65 70

PCT/IB98/01233

257

- (2) INFORMATION FOR SEQ ID NO: 330:
 - (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 58 amino acids
 - (B) TYPE: AMINO ACID
 - (D) TOPOLOGY: LINEAR
 - (ii) MOLECULE TYPE: PROTEIN
 - (vi) ORIGINAL SOURCE:
 - (A) ORGANISM: Homo Sapiens
 - (F) TISSUE TYPE: Colon
 - (ix) FEATURE:
 - (A) NAME/KEY: sig_peptide
 - (B) LOCATION: -22..-1
 - (C) IDENTIFICATION METHOD: Von Heijne matrix
 - (D) OTHER INFORMATION: score 3.5

seg AAPVAAGLGPVIS/RP

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 330:

Met Phe Glu Glu Pro Glu Trp Ala Glu Ala Ala Pro Val Ala Ala Gly
-20 -15 -10

Leu Gly Pro Val Ile Ser Arg Pro Pro Pro Ala Ala Ser Ser Gln Asn
-5 5 10

Lys Gly Ser Lys Arg Arg Gln Leu Leu Ala Thr Leu Arg Ala Leu Glu 15 20 25

Ala Ala Ser Leu Ser Gln His Pro Pro Met
30 35

- (2) INFORMATION FOR SEQ ID NO: 331:
 - (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 30 amino acids
 - (B) TYPE: AMINO ACID
 - (D) TOPOLOGY: LINEAR
 - (ii) MOLECULE TYPE: PROTEIN
 - (vi) ORIGINAL SOURCE:
 - (A) ORGANISM: Homo Sapiens
 - (F) TISSUE TYPE: Lung (cells)
 - (ix) FEATURE:
 - (A) NAME/KEY: sig_peptide
 - (B) LOCATION: -23..-1
 - (C) IDENTIFICATION METHOD: Von Heijne matrix
 - (D) OTHER INFORMATION: score 3.5

seq ALYNIIYVCGIQG/IT

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 331:

258

Met His Ile Tyr Thr Gly Ile Lys Tyr Ile Ala Leu Tyr Asn Ile Ile -20 -15 -10

Tyr Val Cys Gly Ile Gln Gly Ile Thr Phe Met Arg Ala Arg -5 • 1 5